ELEMENTI DI STATISTICA DESCRITTIVA

1. ESEMPI INTRODUTTIVI: TRE SITUAZIONI CHE POSSONO PORTARE AD UNA INDAGINE STATISTICA

Esempio a)

Ogni settembre, all'inizio dell'anno scolastico, il professor Curiosi deve far conoscenza con due classi novelle: la I A e la I B.

A tale scopo, da qualche tempo ormai egli ha preso l'abitudine di somministrare ai ragazzi sempre il medesimo test di ingresso, calcolando poi il punteggio acquisito da ciascun alunno, punteggio che può andare da un minimo di 0 a un massimo di 100.

Quest'anno gli esiti, sui 24 studenti di I A e sui 21 di I B, sono stati i seguenti:

I A (24 allievi)	51 62 42 58 60 68 61 68 64 70 71 60 51 62 41 51 36 47 58 73 37 54 63 65
I B (21 allievi)	45 48 51 63 51 60 29 52 47 41 52 50 56 62 57 70 55 64 59 55 67

Se l'è cavata meglio, nel complesso, la I A o la I B?

Il test conferma o no l'impressione, riportata dal professor Curiosi nel corso delle primissime lezioni, che in una delle due classi ci sia *maggiore omogeneità di rendimento* e nell'altra invece si abbiano parecchi alunni bravi, ma anche parecchi scarsotti?

Se nei 4 anni scolastici precedenti i punteggi erano stati quelli della tabella sottostante:

I A anno scorso	52 58 30 39 61 58 56 48 45 27 40 64 68 50 51 47 52 52 58 39
I B anno scorso	45 77 52 60 75 41 47 71 51 43 59 60 54 55 63 60 35 49 48 65 54 38 48 50 45
I A 2 anni fa	50 57 51 54 56 49 40 54 57 47 60 66 68 40 70 56 72 48 40
I B 2 anni fa	75 70 55 58 68 78 54 50 58 65 71 54 49 44 46 56 65 45 56 50 42 69 41
I A 3 anni fa	54 79 52 60 75 41 47 71 51 43 59 50 54 55 63 60 35 25 26 65 54 25 48
I B 3 anni fa	40 74 59 67 62 69 60 69 57 45 56 62 60 59 79 70 60 65 60 88 40 66 48
I A 4 anni fa	38 55 67 49 57 45 56 55 69 44 35 48 53 61 69 45 67 54 62 72 47 62 52 46 55
I B 4 anni fa	52 56 65 75 55 69 84 70 60 74 76 67 73 74 51 65 55 50 55 73 57 62 23 36 49 42

… questi dati *suffragano* o *non suffragano* la lagnanza, consueta in Sala Insegnanti, che "negli ultimi anni, la preparazione dei ragazzi va sempre più abbassandosi"?

Esempio b)

Negli uffici pubblici di una capitale europea l'assessore competente, dopo alcuni episodi antipatici, ha deciso di testare l'apprezzamento o meno dei cittadini riguardo al lavoro degli impiegati comunali, dando a ogni utente la facoltà di compilare il questionario che segue:

- Nome dell'impiegato …
- Professionalità: (Ottima/Discreta/Sufficiente/Insufficiente/Pessima)
- Cortesia: (Ottima/Discreta/Sufficiente/Insufficiente/Pessima)
- Capacità di gestire i tempi del lavoro: (Ottima/Discreta/Sufficiente/Insufficiente/Pessima)

Come potrà il nostro assessore rappresentare graficamente questi dati in modo da poterli discutere col Sindaco in maniera comoda ed efficace?

Esempio c)

Un sondaggio telefonico sulle intenzioni di voto prima delle elezioni comunali in una città di 84000 abitanti, con 3 candidati sindaco, ha coinvolto 250 persone.

In che misura è attendibile?

2. DUE RIGHE DI STORIA

Sebbene un'attività pratica di carattere statistico si possa, volendo, far risalire persino a tempi che precedono l'invenzione della scrittura, quando già l'uomo tramite tacche su di un bastone era in grado di effettuare conteggi di persone o animali, per i primi **rilevamenti statistici** più "in grande" occorre attendere

- i **Sumeri** (in tavolette del IV-III millennio a.C. sono annotati elenchi di persone e di cose in loro possesso, plausibilmente allo scopo di imporre tributi),
- gli **Egizi** (censimento effettuato intorno al 3000 a.C., anche per valutare quanti operai si potessero impiegare nelle costruzioni faraoniche),
- o la **Cina** del 2200 a.C. circa, in cui una rovinosa inondazione indusse l'imperatore a registrare circa cento milioni di esseri umani suoi sudditi, rilevandone pure il mestiere, dichiaratamente ai fini fiscali.

Presso il **popolo ebreo** vennero effettuati alcuni **censimenti**, come sappiamo dall'Antico Testamento (e anche il Nuovo ne menziona uno, ordinato dai Romani, all'epoca della nascita di Gesù …).

Nella **Roma** antica furono particolarmente frequenti, per ragioni tributarie o militari.

In **epoca medievale e rinascimentale** si ebbero raccolte di dati su persone, terre e beni ad opera, tanto per fare qualche esempio, di Carlo Magno, di Guglielmo il Conquistatore, di Stati come la Repubblica Veneta; e comunque le parrocchie e i monasteri presero l'abitudine di tenere registri di **battesimi, morti, matrimoni e possedimenti**.

L'inglese **John Graunt** (1620-1674) è considerato il primo studioso di Statistica in senso moderno.
Egli raccolse una gran quantità di informazioni cercando di cogliere in esse regolarità e relazioni varie:

- *E' vero che nascono più femmine che maschi?*
- *O che il suicidio è più diffuso nelle persone che fanno determinati mestieri?*
- *E' possibile prevedere l'andamento futuro della numerosità di una popolazione?*

Un amico di Graunt, **William Petty** (1623-1687), introdusse il termine "**aritmetica politica**", per indicare "l'arte di ragionare mediante le cifre sulle cose che hanno attinenza col governo".

Fra i grandi nomi che si occuparono di aritmetica politica citiamo Christiaan **Huygens** (1629-1695).

Il poliedrico **Leibniz** (1646-1716) si interessò anche a concetti quali "vita media" e "vita probabile".

L'astronomo inglese Edmond **Halley**, 1656-1742 (proprio lui, quello della celebre cometa) è considerato il padre della **matematica assicurativa**.

Tra coloro che, nel porre le basi della **Teoria della Probabilità**, apportarono un contributo fondamentale alla Statistica, citiamo Jacob **Bernoulli** (1654-1705), Abraham **de Moivre** (1667-1754) e Thomas **Bayes** (1702-1761).

De Moivre, per inciso, predisse pure il giorno in cui sarebbe morto (27 novembre 1754) in base a un conteggio matematico legato all'aumento progressivo dei propri minuti di sonno … in questo caso, però, più che di scienza si trattò di "fortuna" … o di autosuggestione … va beh, parliamo d'altro!

Adrien-Marie **Legendre** (1752-1833), Karl Friedrich **Gauss** (1777-1855), e Pierre-Simon de **Laplace** (1749-1827) si occuparono, fra l'altro, del "**metodo dei minimi quadrati**".

Al sommo **Gauss** si devono risultati geniali in molteplici settori della matematica, fra cui la **teoria degli errori di misura** (della **curva "normale", o "gaussiana"**, parleremo nel nostro corso).

Thomas Robert **Malthus** (1766-1834) approfondì il tema dell'**accrescimento della popolazione umana in un ambiente dalle risorse limitate**, come quello del pianeta Terra - argomento di estremo interesse nel presente.

Il belga **Quételet** (1796-1874) studiò gli scostamenti degli individui dal modello astratto del cosiddetto "**uomo medio**".
Osservò, fra l'altro, che un più alto **tasso di criminalità** risulta **correlato non tanto alla povertà, quanto alla disuguaglianza** fra le classi sociali.

Francis **Galton** (1822-1911), cugino di Darwin, applicò la Statistica alla genetica, alla teoria dell'evoluzione, alla psicometria; introdusse il termine "**regressione**", e anche quello di "eugenica" o "**eugenetica**" (= come migliorare la specie umana agevolando la riproduzione degli individui con le caratteristiche ottimali). Purtroppo lo stesso termine si legò, qualche decennio dopo, ai deliri nazisti.

Ronald **Fisher** (1890-1962) e Karl **Pearson** (1857-1936) dedicarono il loro ingegno alla "**Statistica Inferenziale**", ossia a quella branca della Statistica che **si propone di "inferire" (dedurre) informazioni su di una intera "popolazione" a partire dallo studio di un "campione" di essa.**

Così il chimico inglese W. S. **Gosset** (1876-1937), dipendente della ditta Guinness produttrice di birra, si pose il problema di come trattare le informazioni provenienti da **campioni piccoli o piccolissimi** e firmò le sue ricerche con lo pseudonimo Student perché la birreria, per salvaguardare i segreti della produzione, faceva divieto ai suoi impiegati di pubblicare qualsivoglia articolo (ne sentirai parlare se un giorno dovessi occuparti della "distribuzione **t di Student**").

3. DI COSA TRATTA LA STATISTICA; STATISTICA DESCRITTIVA E STATISTICA INFERENZIALE

Un'indagine statistica si occupa di un "**FENOMENO COLLETTIVO**"
(cioè di un fenomeno che si presenta in una pluralità di soggetti … sovente, in *tanti* o *tantissimi* soggetti).

ESEMPI DI FENOMENI COLLETTIVI

- La conoscenza delle lingue straniere.
- I gusti musicali.
- L'età alla quale ci si sposa.
- La presenza di animali domestici negli appartamenti.
- La lunghezza delle piste ciclabili nelle città.

Se, ad esempio, noi fossimo interessati al fenomeno
"utilizzo del telefonino da parte degli studenti di una data scuola",
gli studenti di quella scuola sarebbero la nostra "*popolazione statistica*",
ogni singolo studente sarebbe una "*unità statistica*",
e i "*caratteri* " da studiare sarebbero "quanto, quando e come questi ragazzi utilizzano il telefonino".

Un "fenomeno collettivo" viene preso in esame nell'ambito di una data "**POPOLAZIONE STATISTICA**",
della quale si studia una certa caratteristica o "**CARATTERE**" (NOTA)
andando ad analizzare quali sono le "**MODALITÀ**" con cui questo carattere si può manifestare e verificando,
in ciascun elemento (= "**UNITÀ STATISTICA**") della "popolazione", quale di tali modalità è presente,
per desumere da tutto ciò **conteggi, percentuali, "medie", "indici di dispersione", rappresentazioni grafiche.**

NOTA: sovente si studiano, "in parallelo", sulla stessa popolazione statistica, *più* caratteri di uno stesso fenomeno

ESEMPI

<table>
<tr><td>
a) Fenomeno collettivo: le caratteristiche fisiche.
Possibili "popolazioni statistiche":
❑ l'insieme dei residenti in Italia,
❑ o l'insieme dei cittadini di nazionalità italiana,
❑ oppure l'insieme delle donne nate a Stoccolma
Un possibile carattere: il colore degli occhi.
Possibili scelte per le modalità di questo carattere:
❑ marrone ❑ azzurro ❑ verde ❑ grigio
oppure: ❑ marrone ❑ non marrone
E' una "unità statistica" la singola persona
della quale si rileva il colore degli occhi.
</td><td>
b) Fenomeno collettivo: il livello di istruzione.
Possibili "popolazioni statistiche":
❑ l'insieme dei cittadini italiani dai 30 anni ai 40
❑ oppure l'insieme dei Vigili del Fuoco
Un possibile carattere:
l'attestato scolastico più alto conseguito.
Una possibile scelta per le modalità di questo carattere:
❑ licenza elementare
❑ licenza media
❑ diploma di scuola media superiore
❑ laurea
E' una "unità statistica" la singola persona
di cui constatiamo il grado di istruzione.
</td></tr>
<tr><td>
c) Fenomeno collettivo: il costo degli affitti
Una possibile "popolazione statistica":
❑ l'insieme degli appartamenti in affitto in una determinata località
Due possibili caratteri:
❑ la cifra risultante dal contratto d'affitto
❑ oppure la cifra complessiva, compresi gli accordi "in nero"
</td><td>
(problema: come riuscire a rilevare la "vera" cifra?)
Una possibile scelta per le modalità del carattere:
❑ meno di 300 euro mensili
❑ da 300 a 449 euro
❑ da 450 a 599 euro
❑ dai 600 euro in su
E' una "unità statistica" il singolo appartamento
per il quale si va ad annotare il costo del relativo affitto.
</td></tr>
</table>

Come mostra l'esempio c), una "popolazione statistica" *non* deve essere necessariamente un gruppo di *persone*!
Un'ulteriore situazione: se interessasse lo studio del numero di piccoli che le coniglie di un dato allevamento generano nel corso della loro vita, una "unità statistica" sarebbe una coniglia.

OSSERVAZIONI

Una volta deciso il "carattere" di cui ci vogliamo interessare, la scelta delle sue "modalità" non è univoca!
Così come spetta a noi, secondo i nostri interessi o le nostre esigenze, la scelta della "popolazione statistica".

SINONIMI

"Popolazione statistica" = "popolazione" = "COLLETTIVO statistico" = "collettivo" = "universo".
Il numero delle unità statistiche si dice anche "**NUMEROSITÀ**" della popolazione.

LE DIVERSE TIPOLOGIE DI CARATTERI

I “**caratteri**” si distinguono fra

- “**QUALITATIVI**” (= le cui modalità sono espresse da un aggettivo, da un sostantivo o da un avverbio, es. *il colore dei capelli*, *il grado di soddisfazione rispetto a un prodotto*)
- e “**QUANTITATIVI**” (= modalità espresse da un numero, es. *il peso* o *il reddito di una persona*, oppure *il numero di uova prodotte da una gallina in un mese*).

Fra i caratteri “**QUALITATIVI**”, distinguiamo

- quelli “**ORDINATI**”
- e quelli “non ordinati”, o “**SCONNESSI**”.
  Ad es., è “ordinato” il carattere *“livello di istruzione”*. E’ invece “sconnesso” il carattere *“colore dei capelli”*.

Fra i caratteri “**QUANTITATIVI**”, ce ne sono di

- “**DISCRETI**”: quelli che sono descritti da **numeri interi**, come ad es. il “*numero dei figli*” di una donna, o il “*numero di esami già superati*” da un universitario
- e “**CONTINUI**”: quelli espressi, almeno in linea di principio, da un **numero reale** … ma, soprattutto, quelli dei quali interessa *non tanto* il valore preciso, *quanto* il fatto se siano compresi in un dato **intervallo**.
  Ad esempio, l’*“area di una superficie coltivata”;*
  ma anche *il peso o l’altezza individuali, il tempo che si impiega per percorrere una certa distanza, la larghezza delle strade,* sono da considerarsi caratteri “continui”.
  Rifletti: se chiediamo a una persona di darci la sua altezza in cm, e questa ci risponde “171”, vuol dire che la sua altezza rientra nella fascia fra 170,5 e 171,5 e che quella persona ha scelto l’intero 171 in quanto ha valutato questo valore “tondo” come il più vicino alla “vera” misura, che però intera non sarà ...

DEFINIZIONI DI “STATISTICA”; SUA IMPORTANZA; STATISTICA DESCRITTIVA E INFERENZIALE

Possiamo dire che la **statistica** è la disciplina che, **innanzitutto**,
insegna ad esprimere le caratteristiche salienti di un insieme di dati, anche molto vasto, in modo sintetico, con l’aiuto di numeri dotati di valore “riassuntivo” (le “medie”, gli “indici di dispersione”) e con il supporto di rappresentazioni grafiche svariate.

- La parte più elementare della statistica è la cosiddetta “**STATISTICA DESCRITTIVA**”.
  La statistica *descrittiva* **analizza *TUTTE* le unità statistiche della popolazione considerata**.
- La “**STATISTICA INFERENZIALE**” si occupa invece di estrarre dalla popolazione, quando questa è troppo vasta per poter essere studiata nella sua interezza, un sottoinsieme di unità statistiche detto “**campione**”, che verrà esaminato allo scopo di dedurre (= **inferire**) da questa analisi parziale indicazioni che possano valere per la **popolazione intera**, valutando il **grado di attendibilità** di tali indicazioni.

Ballatori, nel 1980, definì la statistica come
“**disciplina che studia i fenomeni collettivi, cioè quei fenomeni per la cui osservazione (descrizione, misura) è necessaria una massa di osservazioni di fenomeni elementari**”.

E Maccacaro, nel 1975, la definì come il “**saper parlare di ciò che non si conosce: o meglio, parlare correttamente di ciò che non si conosce completamente**”.

NEL MONDO CONTEMPORANEO una grandissima mole di dati viene quotidianamente rilevata e messa a disposizione del pubblico. D’altronde l’elaborazione e la rappresentazione grafica dei dati stessi sono al giorno d’oggi enormemente facilitate e rese veloci dall’utilizzo del **computer.**
In questo contesto **LA STATISTICA ASSUME UN’IMPORTANZA COLOSSALE.**
E’ ormai irrinunciabile anche per il profano saper interpretare, per esempio, l’**attendibilità dei sondaggi,** o i **grafici** e gli **indici** che continuamente giornali e TV ci propongono.

E la **scienza** non potrebbe progredire senza **strumenti statistici** in grado di **sintetizzare le informazioni** e di effettuare **previsioni e valutazioni di affidabilità**!
Due esempi soltanto fra i tantissimi possibili:

1) la **teoria degli errori di misura**;
2) nelle sperimentazioni cliniche di un farmaco su un campione di malati, **il** “***p*-value**”, valore che esprime la probabilità che il buon effetto riscontrato sia dovuto al caso, anziché a una reale efficacia del farmaco (di solito vengono considerati significativi per il test di efficacia valori $p < 0{,}05$).

ORIGINE DEL TERMINE “STATISTICA”

Proviene dalla lingua **italiana** (Ghislini, 1589) e precisamente dalla parola “**Stato**”, con riferimento all’utilizzo dei rudimenti di questa disciplina, nel XVI secolo, per studi amministrativi e politici.

4. LE PRIME TRE FASI DI UN'INDAGINE STATISTICA; TERMINOLOGIA

PRIMA FASE: LA SCELTA DEL "FENOMENO COLLETTIVO" DA ANALIZZARE,
vale a dire la scelta del "CARATTERE" che si vuole studiare,
e della "POPOLAZIONE" di cui ci si vuole occupare.

Come abbiamo già accennato, a volte interesserebbe una "popolazione" nella sua interezza, ma per motivi di tempi, di costi, di fattibilità se ne prende solo una parte, un "campione" (basti pensare ai sondaggi elettorali …); certo, si porrà poi il problema di valutare in che misura l'indagine fatta sul campione possa essere rappresentativa a riguardo della popolazione intera … di questo si occupa la "statistica inferenziale".

Si sceglie, per il "carattere", un insieme di "MODALITÀ".

Il carattere "sesso" fra i bambini di una scuola elementare si può manifestare in sole due modalità:
Maschile o Femminile, ma sovente c'è invece una certa **discrezionalità**:
ad esempio, il carattere "soddisfazione dell'utenza rispetto a un certo prodotto" potrebbe essere analizzato

- nelle 3 modalità "Poco soddisfatto/Sufficientemente soddisfatto/Molto soddisfatto",
- oppure nelle 5 modalità: "Per niente/Poco/Sufficientemente/Molto/Moltissimo",
- oppure ancora domandando di esprimere la propria soddisfazione con un voto, che so, da 0 a 10.

Le fasi successive dell'indagine statistica sono finalizzate innanzitutto a stabilire ed annotare in quante, fra le "unità statistiche", si presenta ciascuna delle "modalità".

SECONDA FASE: LA RILEVAZIONE DEI DATI

Si può effettuare:

- con l'**osservazione o misurazione diretta** (ad es. per il colore degli occhi, o per il peso …, o per rilevare le condizioni di salute di un ammalato al quale sia stato somministrato un dato farmaco, …);
- tramite un'**intervista**;
- tramite un **questionario**, che potrà essere:
  I) a **risposta chiusa**; II) a **risposta aperta** (più laborioso, in questo caso, lo spoglio dei dati, e meno facile l'interpretazione delle risposte …); III) "**semistrutturato**" (è un "misto" fra le due tipologie precedenti).

TERZA FASE: LO SPOGLIO DEI DATI

Si conta, per ciascuna delle "modalità" del "carattere",
***quante* fra le "unità statistiche" presentano *quella* modalità.**
Si annotano questi conteggi in una tabella che prenderà il nome di "DISTRIBUZIONE DI FREQUENZA".

Ad esempio: "*Qual è lo sport che pratichi con maggiore divertimento?*" (in una popolazione di 480 giovani)

	Frequenza assoluta	Frequenza relativa (appross.)	Freq. rel. percentuale (appross.)
Calcio	120	0,25	25 %
Pallavolo o Basket	92	0,19	19 %
Footing, Atletica leggera	62	0,13	13 %
Nuoto	57	0,12	12 %
Altro sport	100	0,21	21 %
Nessuno	49	0,10	10 %
TOTALE	**480**	**1,00**	**100 %**

A volte si annota solo la "**FREQUENZA ASSOLUTA**" **di ciascuna modalità**, ossia il
NUMERO DI UNITÀ STATISTICHE NELLE QUALI QUELLA MODALITÀ SI È PRESENTATA.

Altre volte (come abbiamo fatto nella tabella precedente) si va a calcolare anche la "**FREQUENZA RELATIVA**" ossia **il rapporto, il quoziente, fra la frequenza assoluta e il numero totale di unità statistiche**:

$$\boxed{\textbf{frequenza relativa} = \frac{\textbf{frequenza assoluta}}{\textbf{numero delle unità statistiche}}}$$

La freq. relativa esprime *quale parte, quale frazione* delle unità statistiche presenta quella determinata modalità.

Dire che la frequenza relativa della modalità "calcio" è stata $\frac{120}{480} = \frac{1}{4} = 0,25$ significa affermare che ¼ dei giovani interpellati ritiene, fra gli sport praticati, il calcio come il più divertente.

Se la frequenza relativa viene poi moltiplicata per 100, si avrà la "**FREQUENZA PERCENTUALE**" che ci dice **quante unità statistiche su 100 hanno presentato quella modalità**:

$$\boxed{\textbf{frequenza relativa percentuale} = \textbf{frequenza relativa} \cdot \textbf{100} = \frac{\textbf{frequenza assoluta}}{\textbf{numero delle unità statistiche}} \cdot \textbf{100}}\,.$$

Nel nostro esempio, il calcio ha come frequenza percentuale $0,25 \cdot 100 = 25$: lo predilige il 25% degli interpellati.

GLOSSARIO, SINONIMI (felici e meno felici)

Ribadiamo che prende il nome di
DISTRIBUZIONE DI FREQUENZA
L'INSIEME DELLE COPPIE ORDINATE
(MODALITÀ, FREQUENZA ASSOLUTA);

in altre parole,
LA TABELLA CHE A CIASCUNA MODALITÀ ASSOCIA LA SUA FREQUENZA ASSOLUTA,
ossia il numero delle unità statistiche che presentano quella modalità.

Se il carattere è quantitativo, e le sue modalità - espresse in questo caso da valori numerici - sono ripartite per comodità in intervalli, questi intervalli vengono chiamati le "CLASSI DI FREQUENZA".

Ad esempio, le modalità del carattere quantitativo
"estensione S in metri quadrati dell'appartamento in cui si risiede"
potrebbero essere riunite, all'atto di compilare la tabella delle coppie (*modalità, frequenza*),
negli intervalli seguenti:

$S < 40\,m^2$;
$40\,m^2 \le S < 60\,m^2$;
$60\,m^2 \le S < 80\,m^2$;
$80\,m^2 \le S < 100\,m^2$;
$100\,m^2 \le S < 120\,m^2$;
$120\,m^2 \le S < 140\,m^2$;
$S \ge 140\,m^2$

Avremmo allora 7 "classi di frequenza".

Ovviamente, al momento di scegliere le classi di frequenza dobbiamo fare in modo che
- la loro unione dia tutto l'intervallone delle possibili modalità
- e che, prese due qualsiasi di esse, la loro intersezione sia vuota.

Di una distribuzione di frequenza si dice che è:
- **una "SERIE", se il carattere al quale si riferisce è qualitativo**
- **una "SERIAZIONE", se il carattere è quantitativo.**

Una **distribuzione di frequenza** può anche essere chiamata
- "**MUTABILE statistica"** se il carattere di riferimento è **qualitativo**
- "**VARIABILE statistica"** se il carattere di riferimento è **quantitativo.**

Talvolta gli stessi sostantivi "mutabile", "variabile"
vengono impiegati con riferimento al *carattere* più che alla *distribuzione*.

Questa terminologia
si può benissimo ignorare …
dimenticala pure … tuttavia,
te l'ho citata per avvertirti che
consultando un testo o un sito
di statistica, ci si deve rassegnare
a "digerire" parole che a volte
sembrano fatte apposta
per complicare inutilmente le cose.

Avrai osservato che, in statistica, sovente si hanno più possibilità diverse per dare nomi ai concetti.

A dire il vero, alcuni termini danno l'impressione di non essere molto "azzeccati",
o di venire impiegati più per mettere in sudditanza psicologica il lettore che per agevolarlo …

In queste lezioni cercheremo sempre di adottare la terminologia a nostro modesto avviso più chiara,
informando lo studente sui possibili sinonimi (almeno, su alcuni fra i tanti).

E' noto che **PER SEPARARE LA PARTE INTERA DA QUELLA DECIMALE** in un numero
si può utilizzare la *virgola*, o in alternativa il *punto decimale*.
Noi nel nostro corso scegliamo di norma la prima strada - più diffusa in Italia - ma a volte preferiamo invece
la seconda, per motivi vari di opportunità (e un po' anche per abituare il lettore alla doppia possibilità).
♥ **In questo capitolo dedicato alla statistica il separatore sarà la virgola**,
anche perché nel capitolo si invita sovente a servirsi di un *foglio elettronico*
(programma per computer in grado di visualizzare tabelle, effettuare calcoli e tracciare grafici),
e in un foglio elettronico (versione italiana) occorre forzatamente fare uso della virgola, in quanto
un numero scritto con un punto verrebbe interpretato dal programma come finalizzato a indicare un'*ora del giorno*.

Le **FASI SUCCESSIVE** dell'indagine statistica, di cui andremo ad occuparci nel seguito, consistono
- nella **RAPPRESENTAZIONE GRAFICA** dei risultati
- e nella loro **ELABORAZIONE STATISTICA**.

Ma a questo punto ti direi di fermarti per fare qualche facile (e divertente) esercizio.

ESERCIZI sui concetti introduttivi alla Statistica Descrittiva (risposte a pag. 422)

1) Per ognuno dei seguenti fenomeni collettivi, individua una possibile "popolazione" e uno o più possibili "caratteri", poi per ciascun carattere uno o più possibili insiemi di modalità:
   a) le caratteristiche fisiche delle persone b) il lavoro c) le abitudini di spesa
   d) l'appartamento in cui si vive e) la compagnia degli animali domestici

2) Per ciascuno dei seguenti caratteri, stabilisci se è qualitativo *ordinato*, qualitativo *sconnesso*, quantitativo *discreto* o quantitativo *continuo*:
   a) l'età alla quale una donna ha avuto il primo figlio
   b) la squadra di calcio preferita
   c) il numero di sere in cui uno studente esce abitualmente di casa in una settimana
   d) il peso del proprio zainetto nell'entrare a scuola la mattina
   e) il numero di libri presenti nello zainetto di uno studente all'ingresso a scuola
   f) il gradimento di un programma televisivo
   g) la nazionalità degli ospiti di un albergo
   h) il numero di abitanti di un comune

3) Intervista telefonica a 50 persone che hanno risposto "sì" alla domanda "Possiede un gatto?"
   Fra parentesi, il numero di risposte.

L'animale vive in casa o fuori?	□ Solo in casa (14) □ Sia in casa che fuori (24) □ Quasi sempre fuori (12)
Quanto ha speso negli ultimi 30 gg per l'alimentazione del gatto?	□ Meno di 10 euro (25) □ Da 10 € a 20 € (22) □ Più di 20 € (3)
Ha fatto ricorso al veterinario per il gatto negli ultimi 12 mesi?	□ Mai (33) □ Una volta (13) □ Più di una volta (4)

a) Qual è la "popolazione statistica" in questo caso? Quale il "fenomeno collettivo"? Quali i suoi "caratteri"?
b) Riconosci, fra i caratteri studiati, quelli "quantitativi" e quelli "qualitativi"
c) Quali sono le "modalità" scelte per ciascun carattere?
d) Il carattere qualitativo considerato è "ordinato" o "sconnesso"?
e) Per il ricorso al veterinario, determina le frequenze: I) assolute II) relative III) percentuali

4) *Ottimo per un divertente lavoro di gruppo.*
   a) Trascrivi il questionario sottostante al computer con un *word processor* (= programma di *elaborazione testi*), ad esempio *Word* o *OpenOffice Writer*.
   b) Stampa. Fotocopia. Distribuisci.
   c) Raccogli i questionari compilati.
   d) Con un *foglio elettronico* (es. *Excel* o *OpenOffice Calc*) salva gli esiti in un file.

QUESTIONARIO DEL CURIOSONE

Grazie ♥ se vorrai riempire questo questionario, RIGOROSAMENTE ANONIMO!!!

1) La tua altezza, in cm: ………… 2) Il tuo peso, in kg: ………. 3) Il tuo numero di scarpe: …………..
4) Il numero dei tuoi fratelli (escluso te; devono avere la stessa tua mamma e papà naturali): …………..
5) Quanti anni aveva tua mamma quando ha avuto il primo figlio? …………..
6) A che ora vai a letto, di solito, quando il giorno dopo devi andare a scuola? (è ammessa anche la "1/2 ora", es. 22:30) …………..
7) Quanto prendi di "paghetta" mensilmente? Euro …………..
8) Qual è la tua materia preferita? ………… 9) Qual è la materia che trovi più antipatica? ………….
10) Dai un giudizio sulla tua scuola attuale (crocia la lettera corrispondente alla risposta):
    Pulizia, servizi igienici, stato dell'edificio: a) Scarsissima b) Scarsa c) Sufficiente d) Buona e) Ottima
    Preparazione del corpo insegnante: a) Scarsissima b) Scarsa c) Sufficiente d) Buona e) Ottima
    Capacità degli insegnanti di capire i ragazzi: a) Scarsissima b) Scarsa c) Sufficiente d) Buona e) Ottima
11) Quanti telefonini hai posseduto fino ad ora? ………… 12) A che età hai avuto il primo? …………..
13) Qual è il massimo numero di cm che hai saltato in alto in palestra? ………….
14) Sei iscritto a Facebook? ………….. 15) Se sì, quanti "amici" hai? (pressappoco!): …………..

Sei maschio o sei femmina? ………….. Anno di nascita: ………….. Classe: …………..

5) Compila la "distribuzione di frequenze" (assolute, relative e percentuali) per qualcuna delle voci dell'indagine statistica condotta attraverso il "Questionario del curiosone". Esempi:

Altezza, in cm	Frequenza assoluta	Freq. relativa (appross.)	Freq. rel. perc. (appross.)
$150 \le h < 155$	1	0,033	3,3 %
$155 \le h < 160$	3	0,1	10 %
…	…	…	…
Totale	30	1	100 %

Materia più antipatica	Frequenza assoluta	Freq. relativa (appross.)	Freq. rel. perc. (appross.)
matematica	6	0,2	20 %
disegno	5	0,17	17 %
…	…	…	…
Totale	30	1	100 %

6) In un paese di montagna ci sono 40 coppie sposate. Il numero di figli è illustrato dalla seguente tabella.

0	2	1	2	3	1	1	0	2	2	0	0	2	1	2	1	2	1	3	2
2	1	0	2	2	4	0	1	1	2	1	1	1	3	0	1	0	1	1	1

Compila una distribuzione di frequenza, con le frequenze assolute, relative e percentuali.

7) E' stata fatta una indagine di classe:
"Sei soddisfatto della scuola scelta?"
(molto comodo tracciare un'asticella per ogni risposta, poi barrare con un tratto orizzontale i gruppi di 5!)
Ricostruisci il contenuto delle caselle cancellate.

	F. A.	F. R.	F. R. %
Moltissimo			20%
Molto	~~////~~ ~~////~~ //		
Abbastanza	~~////~~		
Poco	~~////~~		
Pochissimo	//		

8) Si sa che 3 delle 4 possibili modalità di un carattere sono state osservate, su di un universo statistico, con frequenze relative 0,35; 0,4; 0,2. Determina la frequenza percentuale della modalità rimanente.

9) VERO O FALSO?
   a) Una volta fissato un carattere, la scelta delle sue modalità è sempre univocamente determinata.
   b) Un carattere è qualitativo quando non ha senso pensare a un ordinamento delle sue modalità.
   c) La statistica inferenziale ha come obiettivo innanzitutto di particolarizzare ad un sottoinsieme, le osservazioni generali riguardanti la popolazione.
   d) In medicina, quanto più il *p-value* è alto, tanto più si può esser persuasi che il farmaco sia "buono".
   e) In una indagine sui costi di una notte in albergo a Roma, 80 euro può essere una unità statistica.

10) Inventa e realizza un'indagine, fra i tuoi compagni di classe, sui seguenti fenomeni collettivi:
   a) il tempo di permanenza davanti ad un monitor, la dipendenza dalla tecnologia
   b) i mezzi di trasporto posseduti in famiglia e quelli abitualmente adoperati
   c) educazione e maleducazione d) cibo per lo stomaco e) cibo per la mente

11) Le medie dei voti in pagella in una classe alla fine del primo quadrimestre.
Raggruppa i dati in classi di frequenza; compila la distribuzione di frequenza

6,75	5,75	6,25	8,13	7,75	6,75	7	7,63	5,5	8,5	6,25	5,63
5	7,88	4,75	8,88	5,25	6,63	6,5	8	7,75	7,5	6,25	5

12) N° di giorni in cui un libro è stato trattenuto in prestito dagli utenti di una biblioteca. Raggruppa i dati in classi di frequenza; compila la distribuz. di frequenza, calcolando anche frequenze relative e percentuali.

7	15	4	10	21	11	9	5	23	28	18	12	15	14	13	22	19	20	7	15	18	12	18
13	12	9	7	9	15	14	10	5	14	16	18	19	21	22	3	16	19	7	26	15	17	12

13) Per le modalità di quali, fra i seguenti caratteri, è opportuna una ripartizione in "classi di frequenza"?
   a) L'ammontare della paghetta settimanale degli adolescenti
   b) Il consumo annuale di acqua di una famiglia
   c) Il numero di quotidiani acquistati da un individuo negli ultimi 30 giorni
   d) La materia più amata dagli studenti e) Il numero di televisori in una casa
   f) Il massimo numero di centimetri realizzati nel salto in alto in palestra
   g) Il voto con cui uno studente è stato promosso in Terza Media (dal 6 al 10)
   h) Il voto con cui uno studente ha conseguito il diploma di scuola secondaria (da 60 a 100 e lode)

14) Come si potrebbe verificare se è attendibile il detto "Donne al volante, pericolo costante"?

15) Con riferimento al paragrafo "due righe di storia", inventa una serie di 7 domande
che possano andar bene per una competizione di classe "*maschi contro femmine*". Parta poi la gara!

16) Ricerca su Internet il significato dei seguenti termini legati alla statistica:
contingenza, scala mobile, indice di ascolto, share, demografia, exit poll, polizza vita.

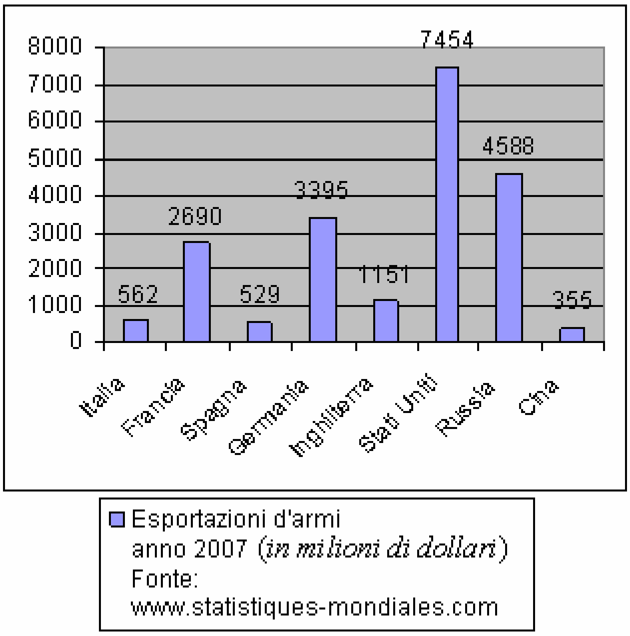
5. RAPPRESENTAZIONI GRAFICHE

Di fronte a un insieme di dati,
indipendentemente dall'intenzione o meno di "fare della statistica" (= calcolare frequenze, medie ecc.),
è sovente assai utile far ricorso a rappresentazioni grafiche,
per agevolare la lettura e l'interpretazione dei dati stessi.

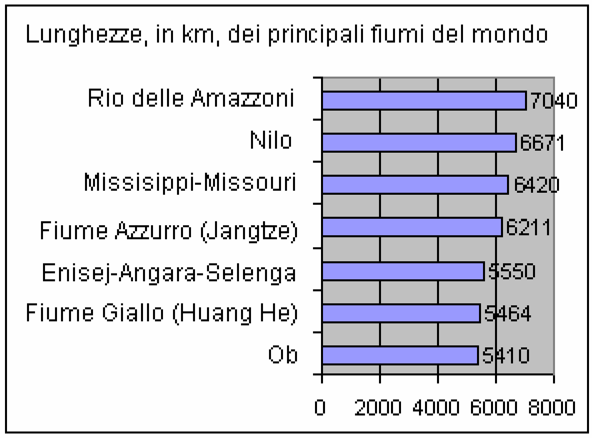
Particolarmente utilizzati sono allo scopo

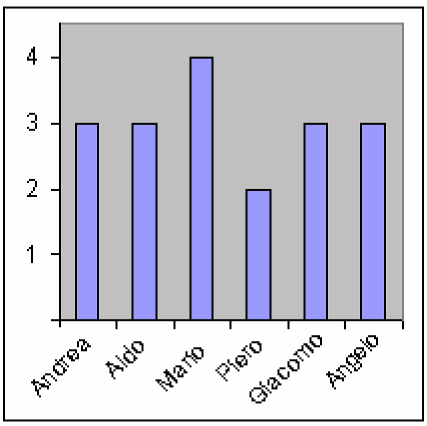
- **I DIAGRAMMI "A BARRE" O "A RETTANGOLI", DETTI ANCHE "ORTOGRAMMI" (rettangoli verticali = "colonne", rettangoli orizzontali = "nastri")**
- **E I GRAFICI CARTESIANI**.

Il loro impiego porta subito ad una visualizzazione molto efficace del fenomeno …
occorre però che I DATI NON SIANO TROPPO NUMEROSI, ALTRIMENTI "CI SI PERDE"!



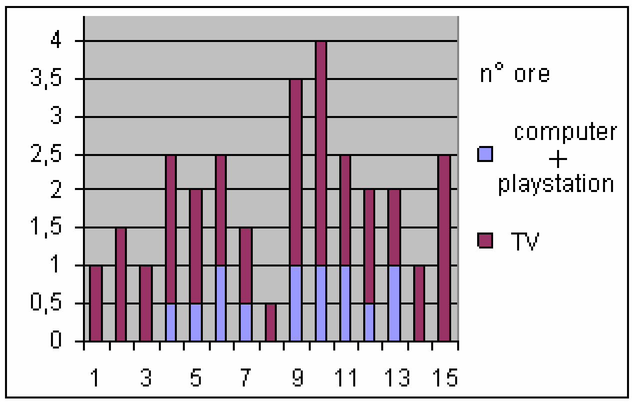
Il diagramma qui a sinistra è "a COLONNE",
quello qui sotto è "a NASTRI".
Entrambi sono diagrammi "a BARRE".
In un foglio elettronico però,
viene detto "diagramma a barre"
solo quello che noi abbiamo chiamato "a nastri".
Che pasticcio, a volte, la terminologia!





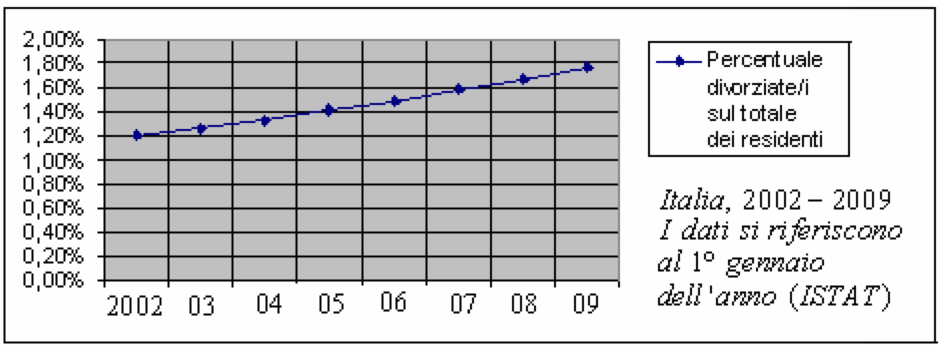
Grado di istruzione delle persone presenti al Bar Sport in un determinato giorno e ora

1 = elementari
2 = medie
3 = diploma
4 = laurea



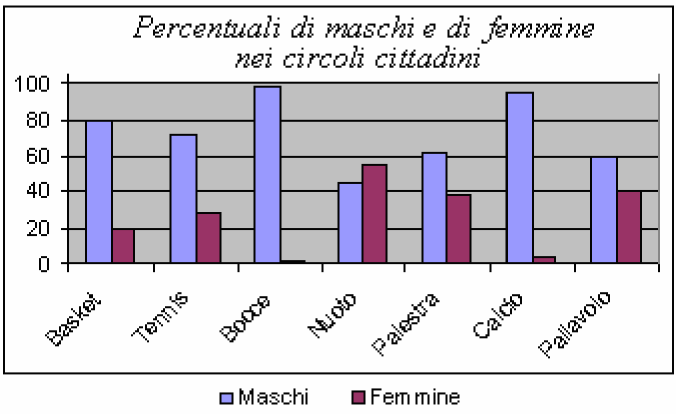
Permanenza media in una giornata davanti a un monitor (per i 15 ragazzi dell'Oratorio S. Giuseppe): barre in pila

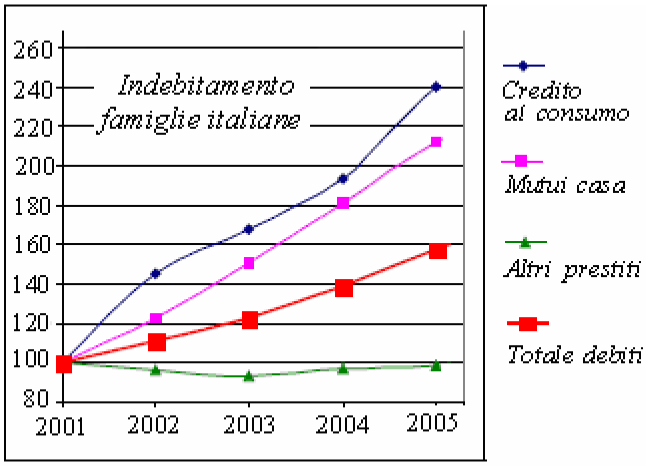
Per una "**SERIE STORICA**", in cui uno stesso dato viene rilevato in tempi successivi, il grafico cartesiano può essere preferibile rispetto al diagramma a rettangoli, perché rende meglio l'idea dell'evolversi del fenomeno:



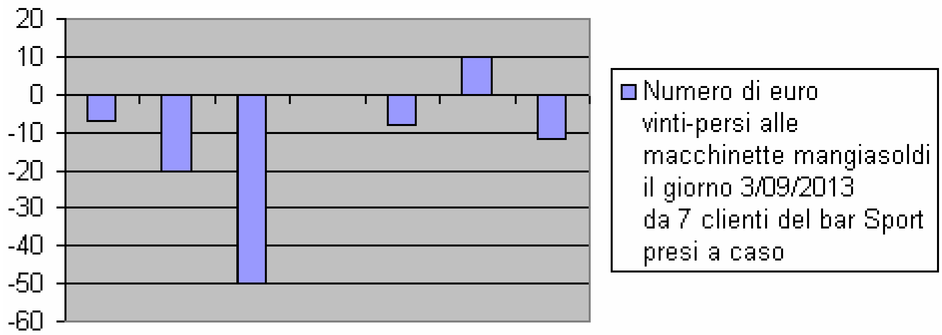
OSSERVAZIONE: le due parole “**DIAGRAMMA**” e “**GRAFICO**” sono in una certa misura intercambiabili, ma in generale sarebbe preferibile limitare l’uso della parola “**grafico**” ai soli casi in cui le quantità che vengono messe in relazione fra loro sono *tutte* espresse da **numeri**, e non da sostantivi/aggettivi/avverbi.

Tanto con un diagramma a rettangoli, quanto con un grafico cartesiano, è possibile anche confrontare *due o più* dati fra loro:



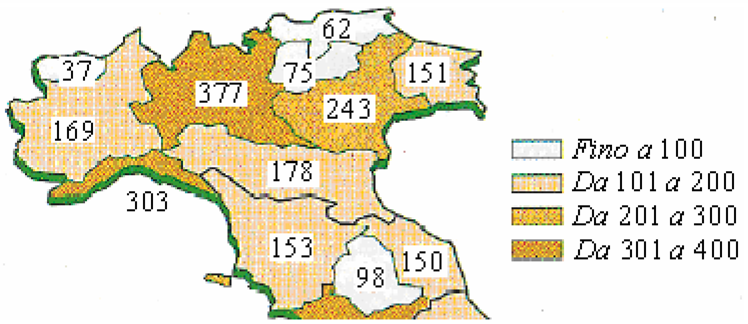


QUANDO I RETTANGOLI SCENDONO AL DI SOTTO DELL’ASSE ORIZZONTALE, esprimono un dato negativo:



In un **CARTOGRAMMA**, le diverse regioni di una cartina geografica vengono colorate con tinte più o meno scure a seconda dell’*intensità* del fenomeno in esame.

Nel cartogramma qui a fianco: intensità della popolazione residente (numero di abitanti per kilometro quadrato, anno 1997).
Fonte: ISTAT



ESERCIZI

1) Rileva le altezze in cm dei tuoi compagni di classe.
Con un **“foglio elettronico”**
(Excel o OpenOffice Calc, ad esempio:
trovi una elementare GUIDA al foglio elettronico nelle pagine successive)
visualizza la situazione.

Tieni presente che

a) **Excel** chiama “Istogrammi” i “diagrammi a colonne” (rettangoli verticali)
e chiama “Barre” i “diagrammi a nastri” (rettangoli orizzontali)

b) **OpenOffice Calc** chiama rispettivamente “Colonna” un “diagramma a colonne” (rettangoli verticali)
e “Barra” un “diagramma a nastri” (rettangoli orizzontali).

Comunque … niente paura!
In un foglio elettronico, la denominazione è subito accompagnata dalla figura, quindi … è facilissimo capire!

2) Con un **“foglio elettronico”** (Excel o OpenOffice Calc, ad esempio)
rappresenta la serie storica della popolazione italiana (in milioni di abitanti)
contenuta nella tabella seguente (dati ISTAT):

1901	1911	1921	1931	1936	1951	1961	1971	1981	1991
33,78	36,92	37,86	41,04	42,4	47,52	50,62	54,14	56,56	56,41

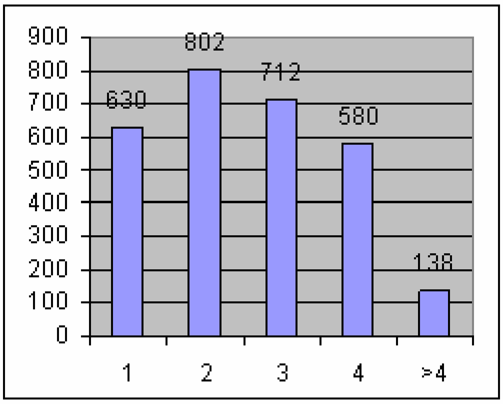
E veniamo ora a esaminare i principali tipi di rappresentazione grafica che si utilizzano più specificatamente per illustrare gli esiti di una vera e propria *indagine statistica*
(c'è un "*collettivo statistico*", o "*popolazione*", e noi andiamo a rilevare
qual è la *frequenza* - assoluta, relativa, o percentuale -
con cui si riscontrano, nella "popolazione", le varie "*modalità*" di un determinato "*carattere*").

DIAGRAMMA A BARRE (= rettangoli verticali o orizzontali)
detto anche "**DIAGRAMMA A RETTANGOLI**" o "**ORTOGRAMMA**".
Se le barre sono VERTICALI, si potrà parlare di "DIAGRAMMA A COLONNE";
se ORIZZONTALI, di "DIAGRAMMA A NASTRI"
(come però abbiamo già fatto notare, **NEI "FOGLI ELETTRONICI" LA TERMINOLOGIA È DIVERSA DA QUESTA. Pazienza, è lo stesso, tanto si capisce ugualmente!**)

E' una figura con rettangoli le cui basi sono fra loro uguali,
e le cui altezze sono proporzionali alle frequenze
(assolute, o relative, o percentuali).

Indagine statistica
sulle 2862 famiglie di un Comune:
numero di componenti
del nucleo familiare
(conviventi nello stesso appartamento)

1	630
2	802
3	712
4	580
>4	138
	2862



I rettangoli si possono eventualmente accostare fra loro.
La figura qui sotto mostra una "tabella composta",
nella quale le modalità di uno stesso carattere sono riferite a più popolazioni.
Notare anche l'aspetto "3D" (= tridimensionale) del diagramma.

"Trovi interessanti i talk show politici in TV"?
(Risposte in percentuale su un campione di 964 intervistati, suddivisi per livello di istruzione)

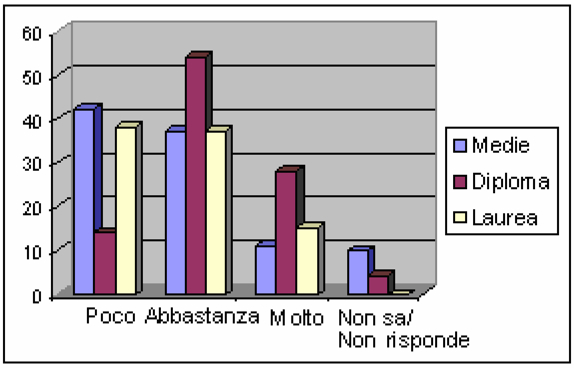
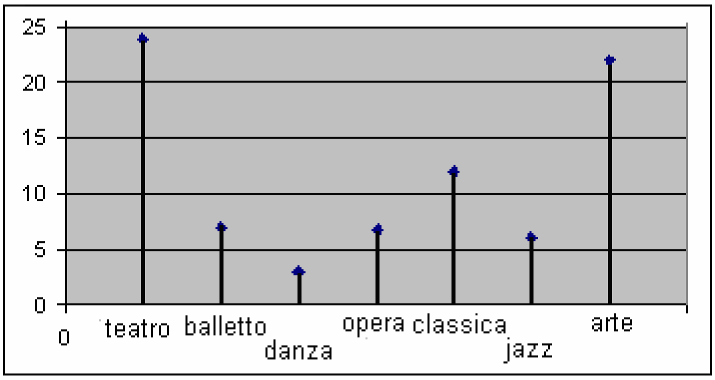


DIAGRAMMA A SEGMENTI (= AD ASTE),
analogo al diagramma a rettangoli, con una delle dimensioni del rettangolo sottilissima.

Partecipazione ad eventi culturali:
percentuale della popolazione
che ha partecipato
ad almeno un evento culturale
del tipo specificato
(teatro, balletto, danza contemporanea,
opera, musica classica,
jazz, galleria d'arte)
nel periodo considerato.



IDEOGRAMMA

Trattoria "Le cascine"

Trattoria "La pergola"

Trattoria "Settefolli"

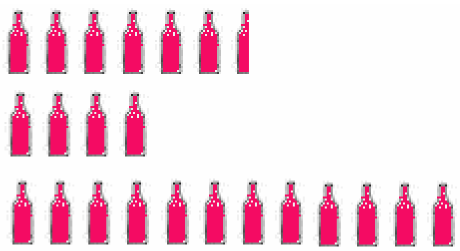
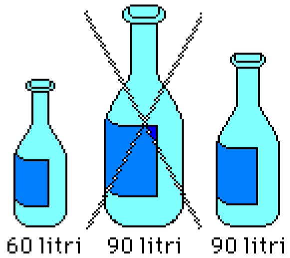


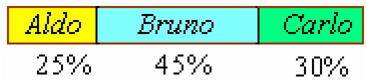
Figure opportune, legate al contesto, vengono disegnate in modo che sia proporzionale al dato da rappresentare o il loro numero, oppure la loro estensione.

Nell'ideogramma qui a fianco:
Il consumo annuo di vino in 3 trattorie
(1 bottiglia = 500 litri)



Attenzione, però:
se è l'*estensione* della figura quella che conta, è facile sbagliare:
ad esempio, nella figura qui a fianco, tratta da http://macosa.dima.unige.it,
la bottiglia centrale dovrebbe avere una volta e mezza il volume della prima
($90 = 60 \cdot 1{,}5$), mentre il realtà, essendo le sue *dimensioni* 1 volta e mezza,
il *volume* è $1{,}5 \cdot 1{,}5 \cdot 1{,}5 = 3{,}375$ volte tanto.
L'ideogramma corretto è la *terza* bottiglia, quella a destra!

DIAGRAMMA A STRISCE
Ottimo per confrontare le parti con il totale



Il contributo di 3 muratori alla piastrellatura di un corridoio

DIAGRAMMA A TORTA (o "**diagramma a settori circolari**")

Un cerchio è suddiviso in tante fette quante sono le modalità del carattere in esame.

L'angolo al centro di una fetta (ossia: di un settore circolare)
è proporzionale alla frequenza, assoluta o relativa o percentuale, della rispettiva modalità.

Come faccio a determinare di quanti gradi x dev'essere una data "fetta"?
Semplice:
Se ad esempio le unità statistiche erano 24, e 10 di esse hanno presentato una certa modalità, allora

$$10 : 24 = x° : 360° \quad \text{da cui} \quad x° = \frac{10 \cdot 360°}{24} = 150°$$

In generale,

$$\begin{matrix}\textit{frequenza}\\ \textit{assoluta}\end{matrix} : \textit{numerosità} = x° : 360° \quad \text{da cui} \quad x° = \frac{\textit{frequenza assoluta}}{\textit{numerosità}} \cdot 360°$$

(ricordiamo che per "numerosità" di una popolazione si intende il numero totale delle unità statistiche).

Si può anche operare (è lo stesso!), per determinare $x°$, con la frequenza relativa:

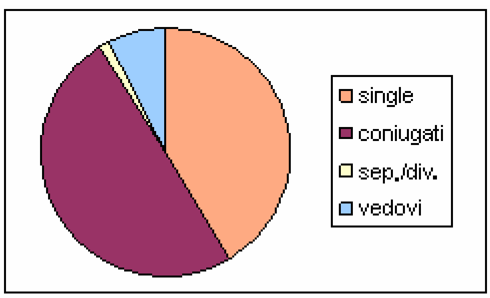
$$\begin{matrix}\textit{frequenza}\\ \textit{relativa}\end{matrix} : \underbrace{\begin{matrix}\textit{somma delle}\\ \textit{frequenze}\\ \textit{relative}\end{matrix}}_{=1} = x° : 360° \quad \text{da cui} \quad x° = \textit{frequenza relativa} \cdot 360°$$

... oppure con la frequenza percentuale:

$$\begin{matrix}\textit{frequenza}\\ \textit{percentuale}\end{matrix} : \underbrace{\begin{matrix}\textit{somma delle}\\ \textit{frequenze}\\ \textit{percentuali}\end{matrix}}_{=100} = x° : 360° \quad \text{da cui} \quad x° = \frac{\textit{frequenza percentuale}}{100} \cdot 360°$$

ESEMPIO: la popolazione (compresi i bambini) di un certo comune per stato civile nel 2013, in migliaia (single, coniugati, separati o divorziati, vedovi). Il fatto che nelle colonne delle frequenze relative e percentuali la somma non sia *esattamente* 1 o 100 è dovuto agli arrotondamenti.

	assoluta	relativa	percentuale	angolo
single	23516	0,414	41,4	149
coniugati	28185	0,496	49,6	179
sep./div.	854	0,015	1,5	5
vedovi	4224	0,074	7,4	27
TOTALE	56779	1,000	100,0	360



ISTOGRAMMA

(utilizzabile, nella versione "per aree" che qui sotto presentiamo, per i caratteri quantitativi continui; tuttavia, quasi sempre il termine "istogramma" viene inteso piuttosto come sinonimo di "diagramma a colonne", il buon vecchio diagramma a colonne con basi delle colonne fra loro uguali. La raffigurazione "per aree" di cui stiamo per occuparci ha infatti sovente più svantaggi che vantaggi)

Un'azienda vuole illustrare la "ripartizione dei suoi dipendenti per classi di età".

Poiché le età di 25 anni, di 45, e di 55, sono normalmente associate a scatti di carriera o comunque appaiono particolarmente adeguate a ripartire i dipendenti in gruppi in qualche modo omogenei (per atteggiamento mentale, per esperienza lavorativa …), viene compilata la tabella seguente
(s'intende, in ogni intervallo, il *primo estremo incluso* e il *secondo escluso*):

da 20 a 25	da 25 a 45	da 45 a 55	da 55 a 68
33	74	18	4

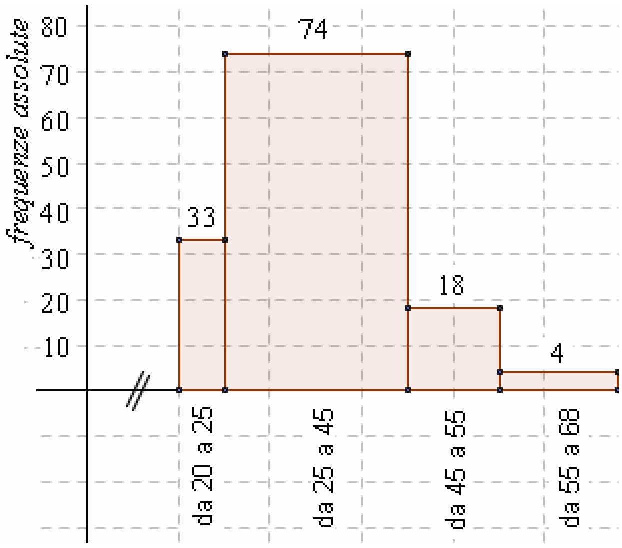
Le "classi" (= gli intervalli) sono fra loro differenti come ampiezza, per cui sembra opportuno che siano pure differenti fra loro (e proporzionali agli intervalli) le suddivisioni dell'asse delle ascisse.

Tuttavia, se a questo punto noi associassimo a ciascuna classe un rettangolo di altezza proporzionale alla frequenza …

… questa rappresentazione potrebbe darci un **"colpo d'occhio" distorto** sulla situazione, per almeno due motivi:

a)
la nostra attenzione è portata spontaneamente a portarsi sull'*area* di ciascun rettangolo, piuttosto che sulla sua altezza … ma da ciò si trarrebbe l'impressione (sbagliata!) che i dipendenti con almeno 45 anni (i 2 intervalli a destra) siano più numerosi di quelli con meno di 25 anni;

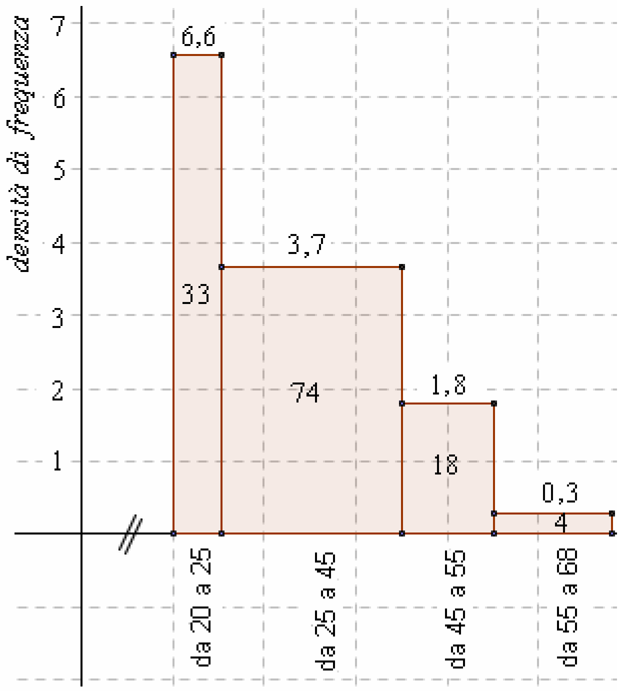
b)
e inoltre, si potrebbe pensare che l'altezza dell'intervallo si riferisca a ogni singolo valore che sta alla base dell'intervallo (quindi, per esempio, che ci siano 33 dipendenti di 20 anni, 33 di 21, 33 di 22 , …)



Viene allora *un'altra idea*.
Su ciascun intervallo si costruisce un **rettangolo** la cui **AREA** sia **proporzionale alla frequenza di quella classe**.

L'*altezza* del rettangolo si potrà perciò ricavare *dividendo la frequenza per l'ampiezza della classe*.

Nel nostro esempio, abbiamo ricavato l'altezza del primo rettangolo a sinistra dividendo la frequenza (che era 33) per l'ampiezza dell'intervallo $(25-20=5)$.
Abbiamo ottenuto $6,6$, quindi il nostro primo rettangolo a sinistra, avendo base 5 e altezza $6,6$, avrà **area** uguale alla frequenza (33) con la quale la modalità "da 20 a 25 anni" si è manifestata nella nostra popolazione statistica (= l'insieme dei dipendenti dell'azienda).



Insomma, **in un istogramma di questo tipo le FREQUENZE *non* sono date dalle altezze dei rettangoli, bensì dalle loro AREE !**

E le *altezze* dei rettangoli vengono anche chiamate "*densità di frequenza*".

QUALCHE OSSERVAZIONE sui diagrammi precedenti.

- La prima è banale: abbiamo utilizzato il simbolo // per indicare il fatto che il segmento in gioco ha una lunghezza che non va d'accordo con le lunghezze degli altri intervalli sull'asse orizzontale: tale segmento non aveva importanza per il nostro diagramma, ed è stato tagliato per guadagnare spazio.
- Inoltre: il riferimento è, in entrambe le figure, "dimetrico", cioè con due diverse unità di misura in orizzontale e in verticale.
  D'altronde, sono addirittura diverse le due grandezze le cui misure vengono riportate sugli assi: in entrambi i diagrammi, l'asse orizzontale riporta intervalli di età, mentre sull'asse *verticale* abbiamo:
  - ♪ nel *primo* dei due diagrammi, una *frequenza assoluta*;
  - ♫ nel *secondo*, una *"densità di frequenza"*, la cui unità di misura ha la dimensione $età^{-1}$
- E' importante, quando si suddividono le modalità in intervalli (= in "classi di frequenza"), specificare con chiarezza se un estremo dell'intervallo è *incluso* oppure è *escluso*.
  Noi lo abbiamo fatto dichiarando "s'intende, in ogni intervallo, il *primo estremo incluso* e il *secondo escluso*".
  A volte si indica l'inclusione o esclusione di un estremo in modo schematico: fra i simboli utilizzati, c'è
  - ⊢ oppure •—∘ per indicare che il 1° estremo è incluso e il 2° è escluso: $x_1 \le x < x_2,\ \ x \in [x_1, x_2)$
  - ⊣ oppure ∘—• per indicare il viceversa: $x_1 < x \le x_2,\ \ x \in (x_1, x_2]$
  - ⊢⊣ o anche solo un trattino $(-)$ oppure •—• per indicare l'inclusione di entrambi gli estremi: $x_1 \le x \le x_2,\ \ x \in [x_1, x_2]$

  Poiché diversi Autori potrebbero effettuare scelte diverse, senza magari esplicitare la loro scelta con chiarezza, occorre sempre, quando si consulta una fonte, cercare di capire come si è regolato quel libro o quel sito.
- Nel caso particolare in cui le ampiezze degli intervalli siano tutte uguali, un istogramma non differisce da quello che noi avevamo denominato "diagramma a barre" o "ortogramma".
- **I "fogli elettronici" EXCEL, OPENOFFICE CALC non fanno istogrammi**, ma solo **diagrammi a barre … … che tuttavia chiamano "istogrammi"! E questa abitudine** a utilizzare il termine "istogramma" per indicare quelli che, per la precisione, andrebbero chiamati "diagrammi a barre", o "ortogrammi", o "diagrammi a rettangoli" **è comunque entrata nell'uso anche nel linguaggio comune.**

  Puoi trovare una brevissima introduzione al "foglio elettronico" nelle pagine seguenti.

Va detto che **GLI ISTOGRAMMI CON INTERVALLI DI UGUALE AMPIEZZA (INDISTINGUIBILI QUINDI DAI "DIAGRAMMI A BARRE") SONO AMPIAMENTE PREFERIBILI**, perché **di più immediata interpretazione**. **IL PREZZO DA PAGARE È CHE LA BASE DEL RETTANGOLO PUÒ NON ESSERE PROPORZIONALE ALL'AMPIEZZA DELLA CLASSE, MA … PAZIENZA!**

Riguardo alle rappresentazioni grafiche, citiamo ancora le "**TABELLE A DOPPIA ENTRATA**": ottime per illustrare la distribuzione di due distinti caratteri su di una stessa popolazione.
Ad esempio: indagine, su 303 famiglie, riguardo a reddito annuo lordo, in euro, e numero di autovetture possedute.

n° auto / Reddito	**0**	**1**	**2 o più**	Totale
<20000	23	27	0	50
20000 ⊢ 40000	5	84	18	107
40000 ⊢ 60000	0	70	29	99
≥60000	0	15	32	47
Totale	28	196	79	303

I TIPI DI RAPPRESENTAZIONE GRAFICA PIU' … "GETTONATI"

Per i caratteri **QUALITATIVI SCONNESSI** i diagrammi più utilizzati sono:
- il diagramma a **torta** (specialmente se le modalità sono poche);
- il diagramma a **colonne** (che Excel chiama *istogramma*) o il diagramma a **nastri** (che Excel chiama *barre*)

Per i caratteri **QUALITATIVI ORDINATI** si utilizzano prevalentemente il diagramma a **colonne** (che Excel chiama *istogramma*) o il diagramma a **nastri** (che Excel chiama *barre*)

Per i caratteri **QUANTITATIVI DISCRETI** si utilizza prevalentemente il diagramma ad **aste** (= **segmenti**)

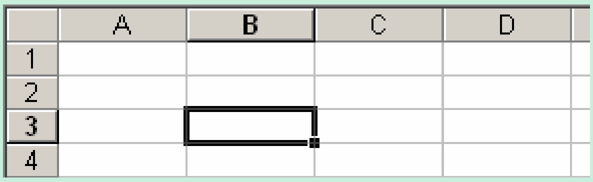
Per i caratteri **QUANTITATIVI CONTINUI** si utilizza prevalentemente il diagramma a **colonne**

Per le **SERIE STORICHE**: grafico **cartesiano,** diagramma a **colonne**

Per le **SERIE GEOGRAFICHE**: **cartogramma**

6. UNA BREVE INTRODUZIONE AL FOGLIO ELETTRONICO

Il “foglio elettronico”



Un “**foglio elettronico**” (o “**foglio di calcolo**”; in Inglese, “**spreadsheet**”)
è un programma per computer che permette di inserire

- **numeri**,
- o **formule**,
- o **scritte** (si dice, in Informatica, “**stringhe**”),

in una griglia di “**celle**” (tipo “**battaglia navale**”), per realizzare
elenchi, tabelle, calcoli e statistiche di vario tipo, e per tracciare **diagrammi e grafici**.

MICROSOFT OFFICE

E’ una **“raccolta” di programmi**, riuniti dalla società produttrice Microsoft in un’unica confezione.

Contiene, nella versione “**Professional**”: **Word+Excel+Powerpoint+Publisher+Access+Outlook.**

Sono comunque in vendita anche “pacchetti” meno costosi, costituiti da un sottoinsieme dei programmi citati.

Tieni poi sempre presente che le **offerte “Education”, riservate a studenti e insegnanti**,
sono molto più economiche delle proposte commerciali “normali”.

❑ Il **foglio elettronico di Microsoft Office** si chiama **Excel** (leggi: *icsèl* o - un po’ “italianizzato” - *ecsèl*)

OPENOFFICE

Una famiglia di programmi simile a Microsoft Office,
composta da programmi che sono di UTILIZZO GRATUITO,
è la famiglia OpenOffice, nata da una iniziativa della *software house* **Sun Microsystems**.

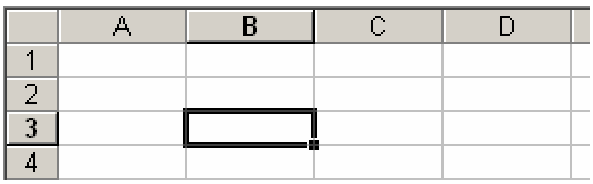
Chiunque può legalmente e liberamente scaricare OpenOffice da Internet accedendo al sito
(in lingua Inglese) www.openoffice.org oppure al sito (in Italiano) http://it.openoffice.org

❑ Il **foglio elettronico di OpenOffice** si chiama **OpenOffice Calc** (l’elaboratore di testi, OpenOffice Writer).

SOMMARIA GUIDA AL FOGLIO ELETTRONICO

Facciamo riferimento per questi brevi cenni a Excel,
ma con OpenOffice Calc il discorso cambia solo in qualche dettaglio.

Se lanciamo il programma,
ci compare un quadro di celle disposte
♪ su righe (1, 2, 3, 4, …)
♫ e su colonne (A, B, C, D, …, Z,
AA, AB, AC, AD, …, AZ,
BA, BB, BC, BD, …).



Ad esempio, qui a fianco, ci siamo posizionati,
cliccando col mouse o adoperando i tasti freccia,
sulla cella B3,
che il foglio elettronico automaticamente ha evidenziato.

Cosa possiamo scrivere, digitando sulla tastiera, in una cella? Possiamo scrivere:

a) un **NUMERO**, **intero o con la virgola** (**OCCORRE ADOPERARE LA VIRGOLA, E NON IL PUNTO, COME SEPARATORE DELLA PARTE INTERA DALLA PARTE DECIMALE**, perché un numero scritto col puntino verrebbe interpretato dal programma come se fosse finalizzato a indicare un’*ora della giornata*, e ciò porterebbe a tutta una serie di esiti sballati)

b) oppure una **SCRITTA** (sequenza di caratteri; in Informatica si dice “**STRINGA**”) di qualsiasi natura

c) o infine una **FORMULA**, **la quale potrà operare sui contenuti di altre celle**.
♥ **UNA FORMULA, PER ESSERE RICONOSCIUTA COME TALE DAL FOGLIO ELETTRONICO, DEVE SEMPRE INIZIARE COL SIMBOLO =**

Facciamo qualche esempio.

Prova a digitare sulla tastiera il numero 58
mentre sei posizionato nella cella A2.
Confermando col tasto “Invio”,
oppure spostandoti col mouse o coi tasti freccia
su di un’altra cella,
il numero 58 diventerà il contenuto di quella cella.

	A	B	C	D
1				
2	58			
3				
4				

Prova a spostarti (mouse, o tasti freccia) sulla cella D1
e digita la scritta
Ciao ragazzi
Essa diventerà il contenuto della cella D1
non appena avrai confermato col tasto Invio
o ti sarai spostato (mouse, tasti freccia) in un’altra cella.

	A	B	C	D
1				Ciao ragazzi
2	58			
3				

Adesso vai sulla cella B2 e scrivi
=A2+1

	A	B	C	D
1				Ciao ragazzi
2	58	=A2+1		
3				

… Bene,
confermando con Invio
o spostandoti in un’altra cella
osserverai che il contenuto di B2 è diventato 59!!!

	A	B	C	D
1				Ciao ragazzi
2	58	59		
3				

Fai qualche altro esperimento …
ad esempio,
inserisci in A3 il numero 7

	A	B	C	D
1				Ciao ragazzi
2	58	59		
3	7			

poi in B3 la formula
=A3*A3
(l’asterisco indica moltiplicazione;
invece la divisione si esprime con la barra /).

	A	B	C	D
1				Ciao ragazzi
2	58	59		
3	7	=A3*A3		

Bene,
dopo la conferma il contenuto di B3 diventerà 49.

	A	B	C	D
1				Ciao ragazzi
2	58	59		
3	7	49		

Adesso posizionati in B4 e digita
= B2 – B3

	A	B	C	D
1				Ciao ragazzi
2	58	59		
3	7	49		
4		=B2-B3		

Naturalmente, dopo la conferma,
il contenuto di B4 diverrà 10.

	A	B	C	D
1				Ciao ragazzi
2	58	59		
3	7	49		
4		10		

Tutto ciò è carinissimo, ed apre la strada a innumerevoli impieghi di straordinaria utilità, soprattutto perché una cella può essere "incollata" su una o più altre celle, e quando il "copia e incolla" viene effettuato a partire da una cella che originariamente conteneva una formula (e ora contiene il numero ottenuto dall'applicazione di tale formula), ciò che viene incollato non è il contenuto visibile della cella, ossia il numero, ma proprio la formula "sottostante" … però la formula che verrà riportata nella cella di destinazione non sarà più esattamente quella originaria, bensì …

Calma, FACCIAMO UN ESEMPIO.

Supponi di essere un commerciante, un artigiano o un imprenditore e di avere una lista di prezzi "al netto di IVA". Cosa vuol dire? Vuol dire che quando farai pagare effettivamente quella merce o quel servizio al cliente, il prezzo non sarà più quello lì, perché dovrai aggiungere una percentuale chiamata IVA (Imposta sul Valore Aggiunto), a carico del cliente stesso.
Per la maggior parte dei beni di mercato, l'IVA era fissata, in Italia, fino all'anno 2010, al 20%.

Ad esempio, un prezzo senza IVA di euro 130 diventava, se "ivato", euro $130+\frac{20}{100}\cdot 130=\frac{120}{100}\cdot 130=156$.

L'IVA è presente in tutti i paesi europei. In Italia è stata a lungo al 20% con l'eccezione dei generi alimentari di prima necessità o dei prodotti di stampa, ivati al 4 %, e di determinati beni e servizi, ivati al 10%.
Nelle altre nazioni si hanno aliquote diverse.

Dopo questa premessa,
immagina di aver stilato con un foglio elettronico
un elenco di prezzi
(celle A2 … A11),
e di voler caricare su di essi l'IVA
(che in questo esempio supponiamo essere del 20 %).

	A	B
1	Senz' IVA	Con IVA
2	130	
3	180	
4	124	
5	120	
6	65	
7	100	
8	82	
9	48	
10	220	
11	30	

Nella cella B2
scriverai dunque
$=A2*120/100$ …

	A	B
1	Senz' IVA	Con IVA
2	130	=A2*120/100
3	180	

… dopodiché,
premendo Invio o spostandoti,
col primo prezzo sarai a posto.

	A	B
1	Senz' IVA	Con IVA
2	130	156
3	180	

Osserva fra l'altro che, confermata la formula nella cella, la formula stessa, ossia il "contenuto concettuale" della cella, viene evidenziato in un'apposita casella in alto …

B2 | fx =A2*120/100

	A	B	C	D
1	Senz' IVA	Con IVA		
2	130	156		
3	180			

A questo punto, si potrebbe temere che facendo un "copia-incolla" della cella B2 sulle celle B3 … B11, la cosa sia destinata a non funzionare in quanto la formula $=A2*120/100$ fa riferimento al contenuto di A2, mentre noi siamo ora interessati a un calcolo del 120 % sui contenuti di A3, A4, …, fino ad A11.

E invece no! Il "miracolo" del foglio elettronico è che,
quando una cella il cui contenuto "concettuale" è una formula viene incollata,

♪ ♥ prima di tutto questo copia-incolla viene applicato *non* al contenuto effettivo *ma* al retrostante contenuto concettuale, appunto.
Insomma, nel nostro esempio, quando incollo la cella B2,
io non incollo il numero 156, bensì la formula "sottostante"

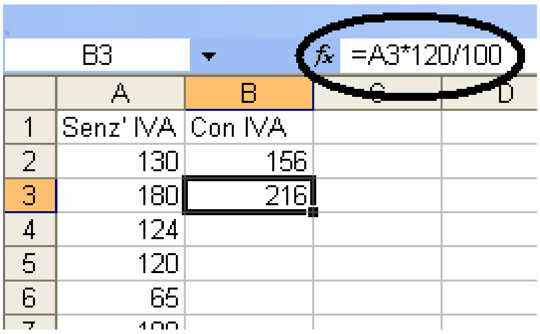
♫ ♥ … e contestualmente, questa formula viene interpretata con "**INDIRIZZAMENTO RELATIVO**" e *non* con "indirizzamento assoluto".
La formula $=A2*120/100$, che abbiamo scritto in B2,
viene interpretata dal foglio elettronico "in senso relativo", "dal punto di vista di B2":

ora, dal punto di vista di B2, $=A2*120/100$ **significa** $=\begin{matrix}\textit{il contenuto}\\ \textit{della cella}\\ \textit{all'immediata}\\ \textit{sinistra}\end{matrix}*120/100$

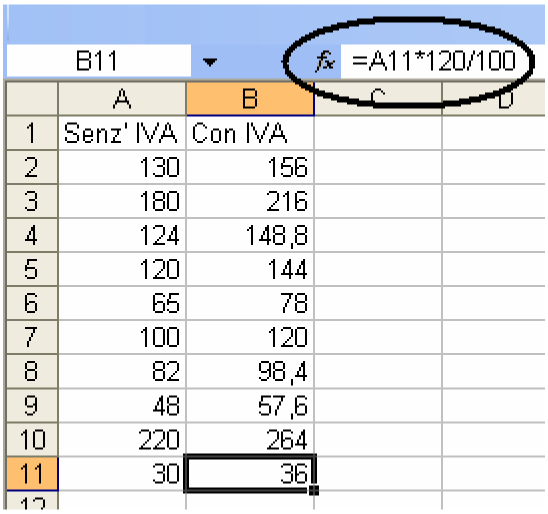
Di conseguenza, quando incolliamo B2 su B3,
la formula che verrà immessa in B3 sarà proprio

$$= \begin{matrix}\textit{il contenuto}\\ \textit{della cella}\\ \textit{all'immediata}\\ \textit{sinistra}\end{matrix} *120/100$$

ossia (siamo in B3)
$$=A3*120/100$$
che fa perfettamente al caso nostro!!!

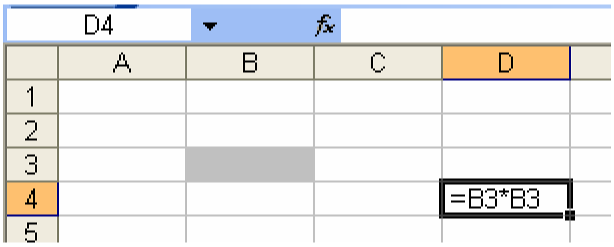


… Ragion per cui,
se noi incolliamo B2
su tutta la serie di celle B3 … B11
(e lo possiamo fare anche "in un colpo solo",
se, dopo aver copiato B2,
prima di incollare
noi trasciniamo il mouse
in modo da selezionare tutte le celle
incolonnate da B3 a B11),
il nostro obiettivo sarà raggiunto.

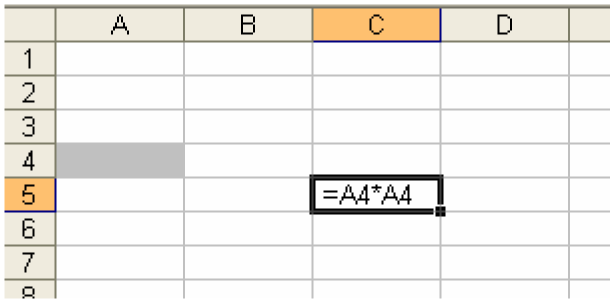


NOTA: *volendo selezionare tutta una zona di celle,*
oltre che trascinare col mouse si può anche
tener premuto il tasto MAIUSC
e poi operare coi tasti freccia.

Un altro esempio: LA FORMULA
$= B3*B3$,
SCRITTA IN D4,
significa, dal punto di vista di D4,
"moltiplicare per sé stesso il numero contenuto
nella cella che sta *2 posti a sinistra e 1 posto in alto*"…

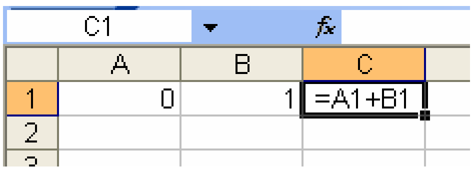


… quindi, SE VENISSE INCOLLATA SU C5,
DIVENTEREBBE
$= A4*A4$

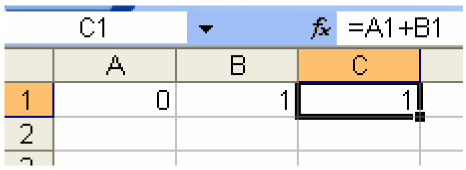


♥ **A volte questo meccanismo dell' "indirizzo relativo" può essere sfruttato per un "effetto domino".**
Supponiamo di voler generare la *successione di Fibonacci*, nella quale
i primi due termini sono 0 e 1, e ciascuno degli altri termini è costruito come somma dei due precedenti:
0 1 1 2 3 5 8 13 21 34 55 89...

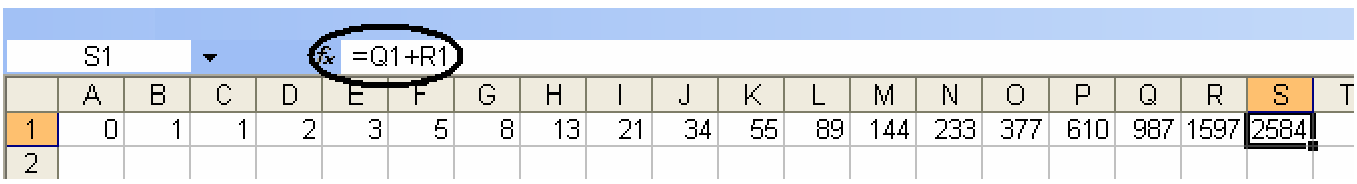
Potremo allora scrivere, ad esempio,
0 in A1, 1 in B1, la formula $=A1+B1$ in C1, …



… ottenendo, tanto per cominciare,



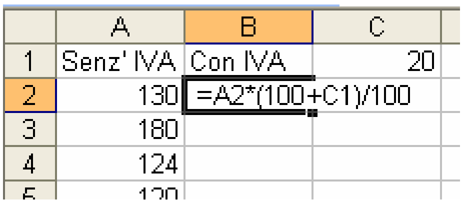
poi incollare C1 sulla fascia D1 E1 F1 ecc., col risultato desiderato.



In certi casi tuttavia l'indirizzamento relativo implicito nelle formule potrebbe esserci d'ostacolo.
L'esempio che segue illustrerà bene questo aspetto, e mostrerà come sia facile fronteggiare la situazione.

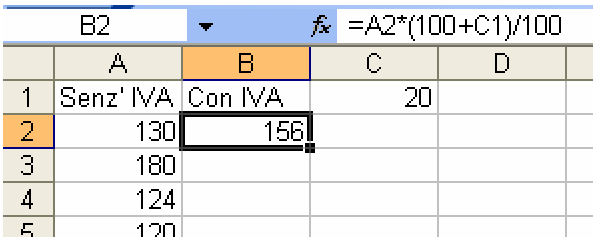
Mettiamoci nuovamente nei panni del commerciante o artigiano alle prese con l'IVA al 20%.
Se il Governo dovesse putacaso abbassarla al 18%, a questo punto il file da lui impostato dovrebbe essere "riprogrammato": ciascuna delle formule contenenti $*120/100$ dovrebbe essere mutata in $*118/100$.
Questo non sarebbe la fine del mondo, ma usualmente i *file* di foglio elettronico sono preparati da un esperto e utilizzati poi da utenti che non sono degli specialisti.

Si potrebbe prevedere una casella in alto, ad esempio C1,
che riporti l'aliquota, poniamo 20,
in modo che anche un utente poco smaliziato
possa ovviare al problema semplicemente scrivendo 18,
o comunque la nuova aliquota, al posto di 20,
in quella cella.
Dopodiché anziché scrivere, in B2, la formula
$= A2*120/100$,
si scriverebbe
$\boxed{= A2*(100+C1)/100}$.



	A	B	C
1	Senz' IVA	Con IVA	20
2	130	=A2*(100+C1)/100	
3	180		
4	124		

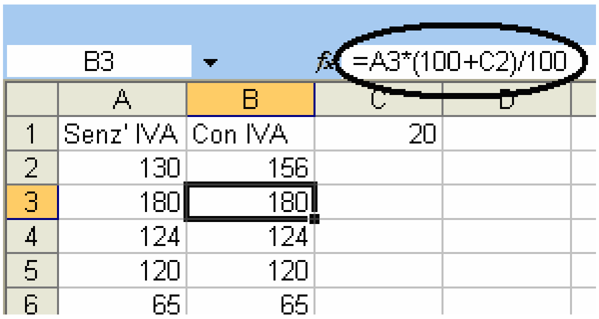
E la cosa funzionerebbe per quanto riguarda B2,
producendo il corretto risultato 156 …



B2 =A2*(100+C1)/100

	A	B	C	D
1	Senz' IVA	Con IVA	20	
2	130	156		
3	180			
4	124			

… però poi, all'atto del copia-incolla su B3, B4 ecc.,
l'indirizzamento relativo farebbe sì che la formula,
incollata ad esempio in B3, divenga
$= A3*(100+C2)/100$
che evidentemente è inservibile
perché in C2 non c'è niente
quindi anziché il desiderato 100+20
il numeratore della frazione
assume il valore 100+0=100.



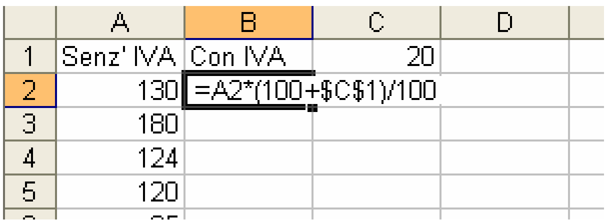
B3 =A3*(100+C2)/100

	A	B	C	D
1	Senz' IVA	Con IVA	20	
2	130	156		
3	180	180		
4	124	124		
5	120	120		
6	65	65		

Fortunatamente i fogli elettronici (Excel, Calc di OpenOffice …) permettono di "inchiodare"
quello che sarebbe un indirizzo relativo, facendolo diventare "assoluto".
Un indirizzo "assoluto" è cristallizzato, immobile, e rimane inalterato anche di fronte a un copia-incolla.

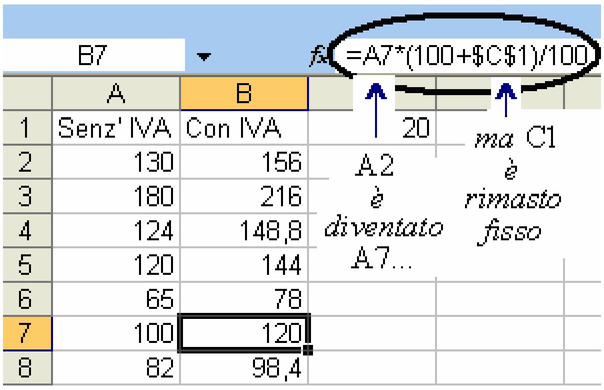
♥ **E' semplicissimo rendere assoluto un indirizzo:
basta usare il "simbolo del dollaro" $.**

Nel nostro caso, dunque,
anziché scrivere in B2 la formula
$\boxed{= A2*(100+C1)/100}$
scriveremo
$\boxed{= A2*(100+\$C\$1)/100}$



	A	B	C	D
1	Senz' IVA	Con IVA	20	
2	130	=A2*(100+$C$1)/100		
3	180			
4	124			
5	120			

Copiando ora la cella B2
e incollandola in B3, B4, …
A2 si muterà in A3, A4, …
ma al contrario C1 rimarrà fisso,
perché bloccato dal "lucchetto" del "dollaro".



B7 =A7*(100+$C$1)/100

	A	B	C
1	Senz' IVA	Con IVA	20
2	130	156	
3	180	216	
4	124	148,8	
5	120	144	
6	65	78	
7	100	120	
8	82	98,4	

E se ora cambiamo il contenuto di C1,
mettendo ad esempio 18 al posto di 20,
non appena confermiamo il 18,
ecco che il foglio elettronico ricalcola
immediatamente e correttamente tutti i valori.

B11 | =A11*(100+$C$1)/100

	A	B	C	D
1	Senz' IVA	Con IVA	18	
2	130	153,4		
3	180	212,4		
4	124	146,32		
5	120	141,6		
6	65	76,7		
7	100	118		
8	82	96,76		
9	48	56,64		
10	220	259,6		
11	30	35,4		

Si può rendere assoluta soltanto la colonna, o soltanto la riga: ad esempio, l'indirizzo contenuto nella formula = $D3 *ha la colonna assoluta e la riga relativa; copiando la formula, e incollandola in un'altra cella, la colonna resterà la D mentre la riga cambierà. Invece nella formula* = E$10 *è assoluta la riga e relativa la colonna.*

Per una conoscenza dettagliata del foglio elettronico rimandiamo ai relativi manuali specifici, o all' "HELP" interno (= la Guida in Linea), attivabile cliccando su "?". Qui ci limitiamo ad alcune osservazioni di carattere generale.

❑ I NUMERI CON LA VIRGOLA; IL MENU *FORMATO*

♥ **Un numero con la virgola va scritto, appunto, utilizzando la virgola e non il punto come separatore. Se infatti un numero viene scritto con un puntino al suo interno o alla fine, viene interpretato dal programma come se indicasse un'*ora della giornata*.**

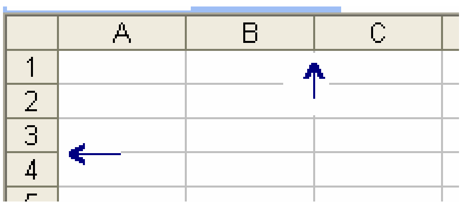
Se in una cella abbiamo per errore scritto un numero col puntino anziché con la virgola, ce ne accorgeremo subito perché il foglio elettronico modificherà automaticamente l'aspetto del numero: ad esempio, un 5.8 viene immediatamente mutato in 5.08 (ore 5 e minuti 8).

C'è poi un altro inconveniente, perché se in quella cella andremo poi a inserire altri numeri, interi o con la virgola che siano, il foglio elettronico li modificherà, in quanto ormai si è "abituato" a interpretare tutti i numeri che vengono immessi in quella cella come indicanti un tempo.
A questo inconveniente si può porre rimedio col menu Formato:
Formato/Celle/Numero e poi si sceglie l'opzione che interessa.

- **I numeri indicanti ore della giornata possono essere anche utilizzati in operazioni aritmetiche, e in questo caso il risultato dell'operazione è coerente con la loro interpretazione.**
  Ad es., 23.00+2.00=1.00
- Il menu Formato permette, fra le tantissime cose, di scegliere il **numero di cifre decimali** alle quali vogliamo che il numero in una determinata cella sia arrotondato

❑ COME RESTRINGERE O ALLARGARE UNA COLONNA O UNA RIGA

Si può trascinare col mouse il margine esterno, nelle posizioni di confine (vedi figura); **oppure**, il che è comodo specialmente se la modifica riguarda tutto un gruppo di celle, **si può selezionare quel gruppo trascinando col mouse e poi andare al menu Formato** per scegliere, ad esempio, Formato/Colonna/Larghezza



❑ SOMMA, MEDIA, ALTRE FUNZIONI "PREDEFINITE"

Per sommare i contenuti di più celle, e porre la somma in E1, si può scrivere, per esempio (in E1)
= SOMMA(A1:D1) .

E1 | =SOMMA(A1:D1)

	A	B	C	D	E	F
1	5	4	2	7	18	

♥ Notare **i "due punti :"** i quali **indicano che si vuole tutta la striscia di celle** fra A1 e D1.

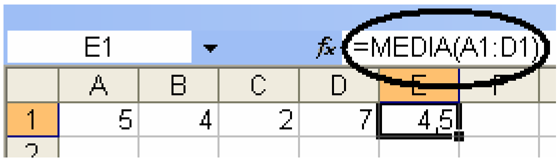
Scrivendo invece col "punto e virgola ;"
= SOMMA(A1; D1)
verrebbero sommati i contenuti soltanto delle celle elencate (le due celle A1 e D1).

E1 | =SOMMA(A1;D1)

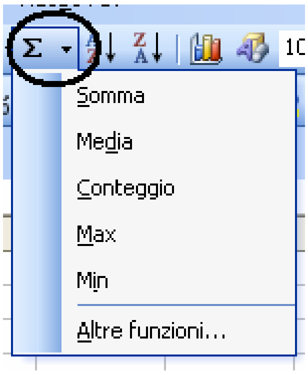
	A	B	C	D	E	F
1	5	4	2	7	12	

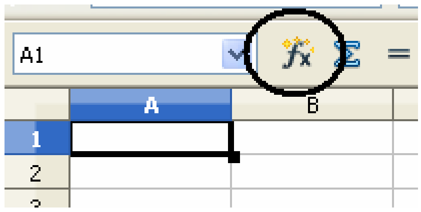
E se si scrivesse invece, ad esempio, = SOMMA(B5:E10) , verrebbero sommati tutti i numeri del **rettangolo di celle** la cui diagonale ha per estremi B5, E10.

Il foglio elettronico è ricchissimo di **funzioni predefinite**.
Solo qualche esempio:
= MEDIA(A1:D1) per la media,
= DEV.ST.POP(A1:D1) per lo scarto quadratico medio
= MAX(A1:D1) per il massimo …



In Excel, per cercare le funzioni disponibili, basta cliccare sull'icona che porta il simbolo di "sommatoria" ...

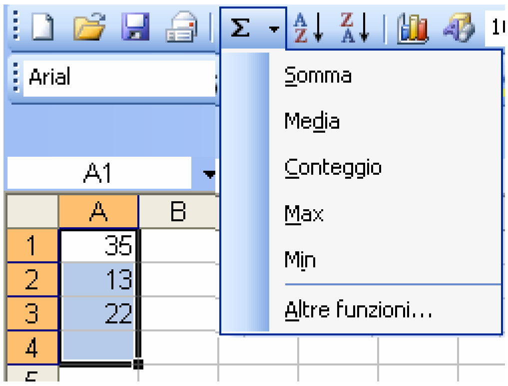




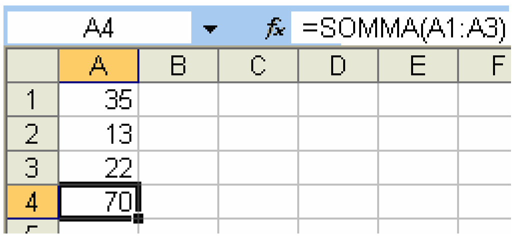
… mentre in Calc di OpenOffice l'icona analoga è quella evidenziata in figura.

❑ SCORCIATOIE

Ci sono anche delle "**scorciatoie**".
Ad esempio, volendo sommare i contenuti delle celle da A1 ad A3 e porre il risultato in A4,
oltre che scrivere, in A4, la formula
= SOMMA (A1:A3)
o la
= SOMMA (A1; A2; A3)
si può, trascinando col mouse, selezionare la fascia di celle da A1 ad A4; poi cliccare su Somma …



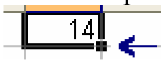
… ed è fatta!



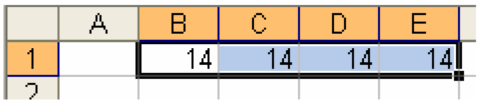
Allo stesso modo per le altre funzioni … tanto per citarne una, la media.

❑ IL QUADRATINO IN BASSO A DESTRA DI UNA CELLA SELEZIONATA

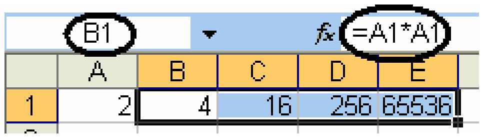
Seleziona ora una cella qualsiasi e vedrai che il foglio elettronico la evidenzia con un bordo marcato che porta **in basso a destra** un **quadratino**



Bene, **trascinando quel quadratino** si può realizzare comodamente il **copia-incolla di quella cella su altre**.

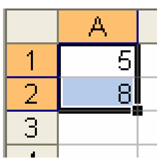


Nella prima delle due figure qui a fianco, il contenuto della cella B1 era 14 ed è stato ricopiato tale e quale in C1, D1, E1.
Nella 2ª figura, la cella B1 conteneva la formula = A1*A1 : l' "effetto domino" ha generato il risultato che puoi vedere.

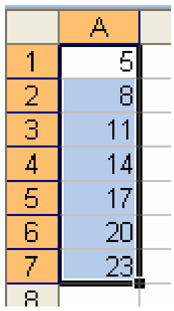


♥ Le "**progressioni aritmetiche**".

Ora scrivi, ad esempio, 5 in A1 e 8 in A2. Seleziona poi la coppia di celle A1, A2 e a questo punto …

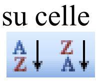


… trascina verso il basso il quadratino. Il foglio elettronico calcolerà la differenza $8-5=3$ e proseguirà automaticamente la sequenza: 5, 8, 11, 14, 17, 20, 23, ecc.!



❑ ORDINARE DATI

E' possibile, ed immediato. Se i dati sono disposti su celle incolonnate una sotto l'altra, basterà cliccare su



per ordinare le righe secondo l'ordine (alfabetico, o numerico) crescente, o decrescente.

❑ LE FUNZIONI LOGICHE

La funzione SE

Un insegnante, dopo l'ultima prova scritta dell'anno scolastico,
essendo soddisfatto per l'interesse e l'educazione che la classe ha sempre dimostrato,
decide di alzare al 4 tutti gli studenti con voto inferiore al 4, e di regalare mezzo punto a tutti gli altri.
Vediamo come potrebbe programmare il foglio elettronico nel quale ha archiviato i voti.

A2 fx =SE(A1<4;4;A1+0,5)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	3	5,5	6	4,5	6,5	8,5	7,5	6	5,5	7	2,5	5,5	8	6,5	9	4	7	6
2	4	6	6,5	5	7	9	8	6,5	6	7,5	4	6	8,5	7	9,5	4,5	7,5	6,5
3																		

La formula
$$= SE(A1 < 4; 4; A1 + 0,5)$$
(che abbiamo scritto in A2 e incollato poi su B2, C2, … R2)
dà al foglio elettronico il comando seguente:

"se il contenuto di A1 è <4, allora in questa cella scrivi 4;
altrimenti, scrivi in questa cella il contenuto di A1 aumentato di 0,5"

La sintassi di una SE è dunque:
$$SE(\mathit{condizione};\ \mathit{se\_vera};\ \mathit{se\_falsa})$$

C'è anche la possibilità di omettere $\mathit{se\_falsa}$.
Puoi provarci, e vedere come si comporta in questo caso il foglio elettronico.
Ricorri comunque, per ogni dubbio, alla Guida in Linea che si attiva cliccando su ?
E' possibile anche annidare più SE uno dentro l'altro.

Le funzioni E, O, NON

- ❑ $E(\mathit{condizione}1;\ \mathit{condizione}2;\ ...)$ restituisce
  VERO se tutte le condizioni sono vere; FALSO se una o più di esse è falsa
- ❑ $O(\mathit{condizione}1;\ \mathit{condizione}2;\ ...)$ restituisce
  VERO se almeno una delle condizioni è vera; FALSO se sono tutte false
- ❑ $NON(\mathit{condizione})$ restituisce VERO se *condizione* è falsa, FALSO se *condizione* è vera.

In questo esempio,
abbiamo scritto in C1 la formula
$= O(A1 > 10; B1 > 10)$
che abbiamo poi copiato
e incollato su C2, C3, C4

C1 fx =O(A1>10;B1>10)

	A	B	C	D
1	5	3	FALSO	
2	12	4	VERO	
3	7	25	VERO	
4	32	14	VERO	
5				

Ecco qui sotto un foglio elettronico che provvede a selezionare i ragazzi vincitori di una borsa di studio.
L'Istituto ha deciso di premiare gli studenti che al termine della classe Seconda
hanno avuto una pagella con media dei voti maggiore o uguale a 9.

Osserviamo per inciso che

- ♪ *le due stringhe SI' e NO desiderate vanno scritte, nella formula, tra virgolette*
- ♫ *le stringhe e i valori appaiono centrati rispetto alla colonna, perché si erano preventivamente selezionate le celle per poi cliccare su*

E2 fx =SE(E(C2=2;D2>=9);"SI'";"NO")

	A	B	C	D	E	F
1	Cognome	Nome	Classe	Media	Borsa?	
2	Abate	Tiziana	4	8	NO	
3	Accursio	Carlo	2	8,3	NO	
4	Adinolfi	Luca	4	6,5	NO	
5	Aggio	Paola	3	7	NO	
6	Altieri	Elisa	2	9	SI'	
7	Amoroso	Katia	1	6,5	NO	
8	Antonazzo	Roberto	3	7,5	NO	
9	Aprile	Loredana	3	7	NO	

- DIAGRAMMI E GRAFICI

Se abbiamo utilizzato un foglio elettronico per realizzare una tabella di dati, (come quella della figura qui a fianco, che si riferisce alle vendite settimanali di un emporio musicale), il foglio stesso ci offrirà la possibilità di tradurre quella tabella in un diagramma, con passaggi molto intuitivi.

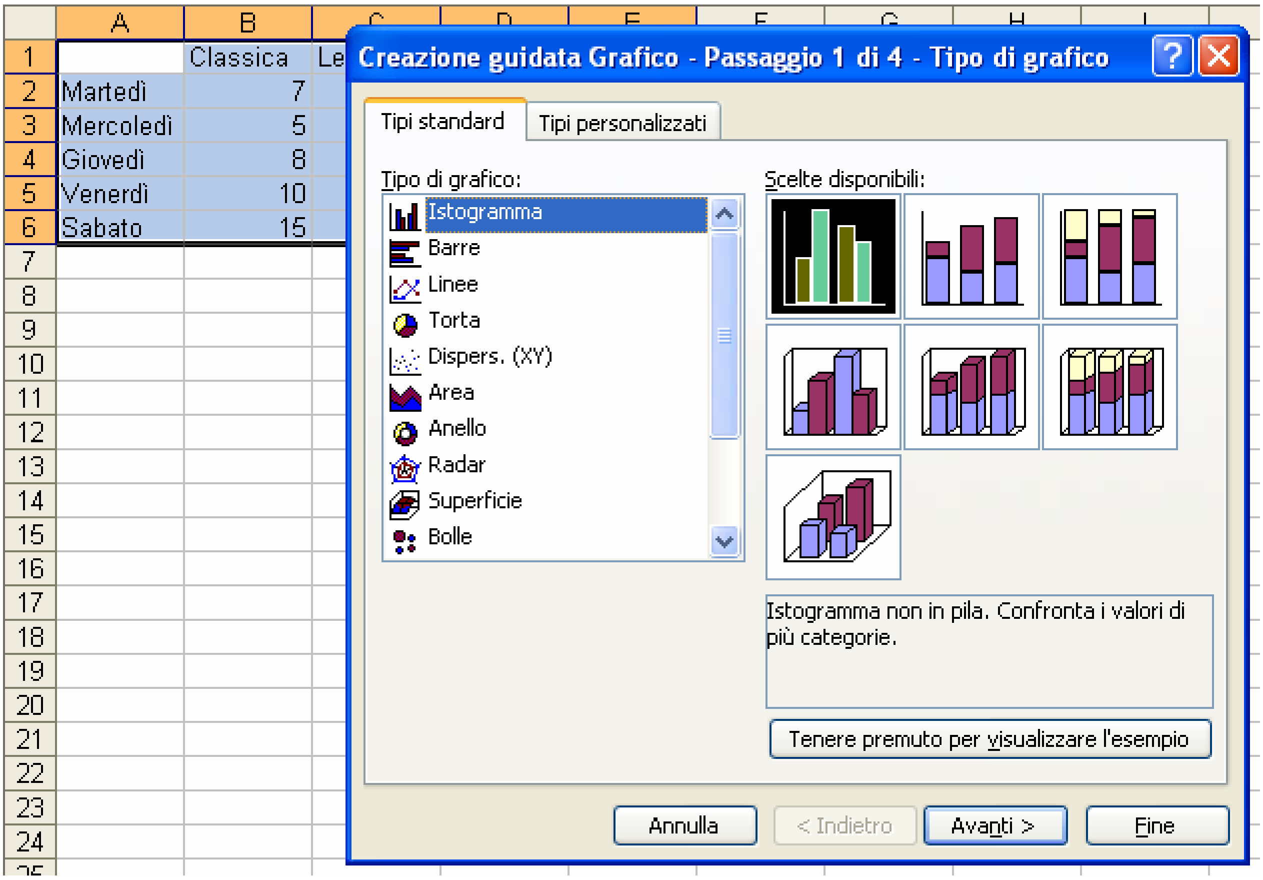
	A	B	C	D	E
1		Classica	Leggera	Rock	Jazz
2	Martedì	7	24	40	3
3	Mercoledì	5	20	35	5
4	Giovedì	8	32	45	2
5	Venerdì	10	36	52	5
6	Sabato	15	44	60	7

Selezioniamo, trascinando col mouse, la tabella, e clicchiamo sull'icona

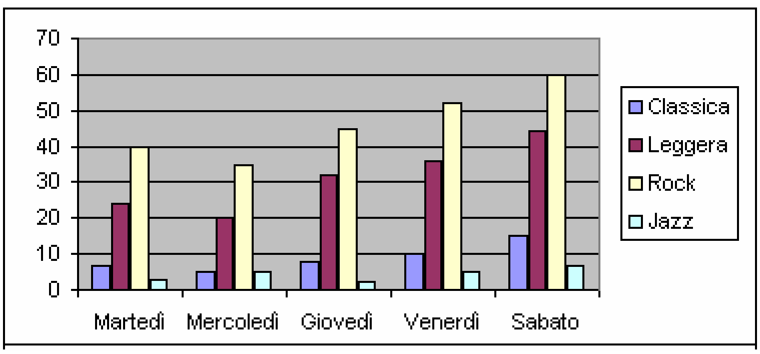
(se non compare, apriamo il menu Inserisci)

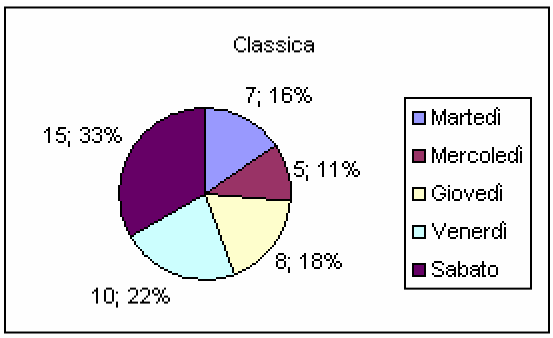
	A	B	C	D	E
1		Classica	Leggera	Rock	Jazz
2	Martedì	7	24	40	3
3	Mercoledì	5	20	35	5
4	Giovedì	8	32	45	2
5	Venerdì	10	36	52	5
6	Sabato	15	44	60	7

Ci troveremo di fronte una finestra tipo la seguente



a partire dalla quale, a colpi di clic, potremo scegliere il tipo di grafico, decidere se includervi valori o percentuali, e completare il diagramma, **controllando passo dopo passo se quello che "esce" è coerente coi nostri desideri**. La pratica diretta ci permetterà di imparare la terminologia e le tante opzioni disponibili.





- CONTA.SE

La formula $=$ CONTA.SE(A1:A20; "<100") , digitata in una cella,
restituisce il numero delle celle, nella fascia da A1 ad A20, il cui contenuto è <100.
Notare le virgolette nella sintassi.

Purtroppo la formula "impazzisce" se tentiamo di inserire una condizione *composta*, tramite una E o una O.

Ma allora, come si potrà procedere, se ad esempio si vogliono contare le celle, nel rettangolo A1:G8, il cui contenuto è compreso fra 15 e 30 (>15 e <30)?
Beh, si ricorrerà ad un "trucco", scrivendo

= CONTA.SE(A1:G8; ">15")– CONTA.SE(A1:G8; ">=30")

Nell'esempio dato, gli estremi 15 e 30 dell'intervallo avrebbero potuto anche essere inseriti in due apposite celle, poniamo ad esempio A11 e A12.
Allora la formula avrebbe dovuto essere riscritta come

= CONTA.SE(A1:G8; ">"&A11)– CONTA.SE(A1:G8; ">="&A12)

- NUMERI CASUALI, O MEGLIO: PSEUDOCASUALI

Possiamo pure ordinare a un foglio elettronico di generare numeri *casuali*, o meglio "*pseudo*casuali": essi infatti hanno l'*apparenza* della casualità, ma in realtà non sono realmente casuali in quanto sono costruiti tramite un algoritmo a partire da un valore iniziale, detto "seme", quello *sì* – ma *solo quello* – da ritenersi casuale (si tratta, di norma, del numero di secondi trascorsi da una certa data del passato).

Digitando in una cella

= CASUALE()   [notare la coppia di parentesi senza niente all'interno!]

si genera, in quella cella, un numero casuale con la virgola x che può andare da 0 (incluso) a 1 (escluso):
$0 \le x < 1$

Questo numero cambierà ogniqualvolta nel foglio elettronico un dato verrà inserito, o cancellato (o anche semplicemente se si preme, posizionati in una cella vuota, il tasto CANC; oppure ancora, premendo il tasto-funzione F9 in alto sulla tastiera); come pure, ad ogni riapertura del file.

- E se volessimo un numero casuale fra 0 (compreso) e 6 (escluso)? Beh, ci basterebbe scrivere
  $= \text{CASUALE}()*6$
- E fra 1 (compreso) e 15 (escluso)?
  $= \text{CASUALE}()*14+1$
- E se volessimo simulare il lancio di un dado, quindi ci servisse un numero INTERO casuale fra 1 e 6? In questo caso potremmo ricorrere a una combinazione fra la funzione CASUALE e la funzione INT.
  INT tronca un numero all'intero più vicino per difetto, quindi, ad esempio, $\text{INT}(3,8)=3$
  Allora la formula
  $= \text{INT}(\text{CASUALE}()*6+1)$
  ci fornirà per l'appunto un intero che potrà valere, con ugual probabilità, 1, 2, 3, 4, 5 o 6.
  Infatti, $= \text{CASUALE}()*6$ genera un numero con la virgola che può andare da 0 (compreso) a 6 (escluso); aggiungendo 1 si ottiene un numero con la virgola che può andare da 1 (compreso) a 7 (escluso); dopodiché la funzione INT, troncando il numero ottenuto, lo fa diventare un intero compreso fra 1 e 6.
- Analogamente, il lancio di una moneta potrà essere simulato da
  $= \text{INT}(\text{CASUALE}()*2)$
  Il risultato dell'applicazione della formula potrà essere il numero 0, oppure il numero 1: bene, "0" potrà essere interpretato come "Testa" e "1" come "Croce", o viceversa.

ESERCIZI sul foglio elettronico

1) *Gli anni di nascita delle persone residenti in un paesino di montagna decedute nell'anno 2009.*

Programma il foglio elettronico in modo che calcoli l'età (approssimativa) in cui sono morte e la media delle età raggiunte.

In pratica:
l'utente inserisce i dati nella colonna A;
il foglio elettronico
riempie automaticamente
la colonna B, compresa la cella B12.

	A	B
1	Anno nascita	Età morte
2	1911	98
3	1920	
4	1919	
5	1922	
6	1923	
7	1915	
8	1944	
9	1930	
10	1928	
11	1920	
12		85,8

2) In C1 il commerciante inserisce la *percentuale di sconto* da concedere, e in colonna A i *prezzi da scontare*.
Occhio alla questione degli *indirizzi relativi e assoluti*!

	A	B	C
1	Prezzo	Scontato	20
2	100	80	
3	200	160	
4	140	112	
5	158	126,4	

3) Ogni studente ha avuto 5 voti. Il foglio elettronico deve calcolare la *media*, e pure individuare la *differenza fra il voto massimo e il voto minimo* (funzioni *MEDIA*, *MIN*, *MAX*, delle quali puoi cercare le caratteristiche cliccando su un'icona come Σ o *fx*)

	A	B	C	D	E	F	G	H
1							Media	Diff
2	Amato	6	7	6,5	8	7	6,9	2
3	Baldi	6	4	6	7	5	5,6	3
4	Carli	6,5	8	6,5	8	7	7,2	1,5

4) Imposta il foglio elettronico in modo che l'utente possa inserire i valori in A2, B2, … , F2 e il foglio calcoli automaticamente le *somme parziali, giorno per giorno*.

	A	B	C	D	E	F
1	Lun	Mar	Mer	Gio	Ven	Sab
2	125	232	167	283	352	420
3	125	357	524	807	1159	1579

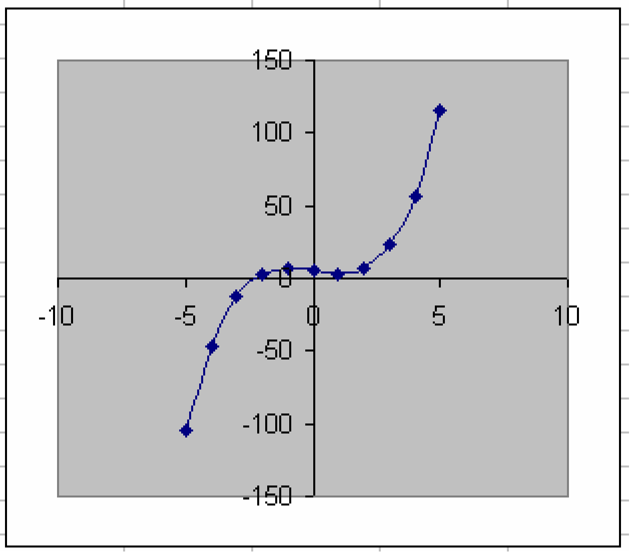
5) a) Calcolo delle potenze successive di 2, scrivendo *una sola formula* da sottoporre poi al copia-incolla:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	2	4	8	16	32	64	128	256	512	1024	2048	4096	8192	16384	32768

b) Utilizza poi il foglio così impostato per verificare che la somma $\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\frac{1}{32}+\frac{1}{64}+\frac{1}{128}+...$ si avvicina, come valore, a 1, quanto più è alto il numero degli addendi che si prendono.

6) Calcolo del *valore di un polinomio di 3° grado*, per vari valori di *x*. *Grafico*. [Indirizzi relativi e assoluti].

	A	B	C	D
1	**VALORE DI UN POLINOMIO DI 3° GRADO ax^3+bx^2+cx+d**			
2				
3	**a**	**b**	**c**	**d**
4	1	0	-3	5
5				
6	**x**	**val. pol.**		
7	-5	-105		
8	-4	-47		
9	-3	-13		
10	-2	3		
11	-1	7		
12	0	5		
13	1	3		
14	2	7		
15	3	23		
16	4	57		
17	5	115		



7) E' bellissima ♥ la *formula (di Leibniz)*:

$$\frac{1}{1}-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\frac{1}{9}-\frac{1}{11}+\frac{1}{13}-...=\frac{\pi}{4}.$$

Essa significa che, con la somma algebrica di tantissimi addendi come quelli a 1° membro, ci avvicineremo a *pi*/4.

Sapresti programmare il foglio elettronico in modo che fornisca una approssimazione di *pi*?

♪ Il simbolo *pi* è sovente utilizzato al posto della lettera greca π per indicare il numero 3,14159265358979 … che interviene nello studio della circonferenza e del cerchio.

♫ Nel foglio elettronico la funzione che restituisce *pi* (o meglio, una sua approssimazione) è PI.GRECO()

	A	B	C	D	E
1	1	1	1	1	
2	3	0,333333333	-1	-0,333333333	
3	5	0,2	1	0,2	
4	7	0,142857143	-1	-0,142857143	
5	9	0,111111111	1	0,111111111	
6	11	0,090909091	-1	-0,090909091	
7	13	0,076923077	1	0,076923077	
8	15	0,066666667	-1	-0,066666667	
9	17	0,058823529	1	0,058823529	
10	19	0,052631579	-1	-0,052631579	
11	21	0,047619048	1	0,047619048	
12	23	0,043478261	-1	-0,043478261	
13	25	0,04	1	0,04	
14	27	0,037037037	-1	-0,037037037	
15	29	0,034482759	1	0,034482759	
16	31	0,032258065	-1	-0,032258065	
17				0,769788349	3,079153394
18				pi/4	pi

8) Questo esercizio è finalizzato a prender confidenza con gli indirizzi “misti”, metà relativi e metà assoluti.
Come riempire la *tabella di moltiplicazione* sottostante inserendo UNA SOLA FORMULA
della quale si farà poi il copia-incolla?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	*****	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	
2	**1**	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
3	**2**	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	
4	**3**	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	
5	**4**	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	
6	**5**	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	
7	**6**	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	
8	**7**	7	14	21	28	35	42	49	56	63	70	77	84	91	98	105	
9	**8**	8	16	24	32	40	48	56	64	72	80	88	96	104	112	120	
10	**9**	9	18	27	36	45	54	63	72	81	90	99	108	117	126	135	
11	**10**	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	
12	**11**	11	22	33	44	55	66	77	88	99	110	121	132	143	154	165	
13	**12**	12	24	36	48	60	72	84	96	108	120	132	144	156	168	180	
14	**13**	13	26	39	52	65	78	91	104	117	130	143	156	169	182	195	
15	**14**	14	28	42	56	70	84	98	112	126	140	154	168	182	196	210	
16	**15**	15	30	45	60	75	90	105	120	135	150	165	180	195	210	225	
17																	

9) Hai un fratellino o una sorellina che fa le elementari e sta imparando le *tabelline*? No? Uhm … fa lo stesso.
Programma il foglio elettronico in modo che in prima e seconda colonna
l’istruttore possa inserire coppie di numeri da lui scelti fra 1 e 10,
in terza colonna il bambino scriva quello che ritiene essere il risultato della loro moltiplicazione,
e in quarta colonna esca OK o NO a seconda dei casi. Prevedi:

- 10 operazioni,
- conteggio automatico delle risposte esatte,
- un complimento al bambino se raggiunge o supera il punteggio minimo posto in F1.

NOTA - *Si può anche rendere casuale la comparsa dei numeri nelle colonne A e B.*
L’unico inconveniente è che quando il bambino risponde … Provaci!

	A	B	C	D	E	F
1			Risultato della "PER"?			7
2	9	10	90	OK		punt. min.
3	3	5	15	OK		
4	7	8	46	NO		
5	4	2	8	OK		
6	3	1	3	OK		
7	2	7	14	OK		
8	4	8	32	OK		
9	5	9	45	OK		
10	5	3	15	OK		
11	6	2	12	OK		
12				9	numero risposte esatte	
13				Bravissimo!		

10) Le celle del rettangolo da A1 ad H8
vengono riempite
con *numeri con la virgola casuali*
da 0 fino a 1 escluso;
si calcola poi
la media di questi 64 numeri.

	A	B	C	D	E	F	G	H
1	0,9016	0,3271	0,3338	0,5098	0,904	0,7361	0,8273	0,8084
2	0,0538	0,9334	0,3281	0,0716	0,8896	0,5204	0,5912	0,2331
3	0,1484	0,7137	0,7485	0,2121	0,0551	0,6182	0,7443	0,056
4	0,3504	0,8705	0,6363	0,8288	0,4879	0,5767	0,0072	0,1711
5	0,8931	0,4227	0,439	0,7296	0,4546	0,8107	0,1633	0,1225
6	0,3052	0,7595	0,1972	0,5094	0,2173	0,6209	0,6785	0,2472
7	0,0696	0,9051	0,4714	0,772	0,0169	0,3088	0,0934	0,7147
8	0,1328	0,4445	0,9574	0,0248	0,887	0,8116	0,4552	0,8249
9	**0,4946**	media						

7. ESERCIZI sulle rappresentazioni grafiche (risposte a pag. 422)

1) La popolazione mondiale, in milioni di abitanti, nel 2009 (dal sito www.statistiques-mondiales.com).

Africa	996
America	923
Asia	4228
Europa	588
Oceania	35
Totale	6776

a) Stabilisci, senza accendere il computer, di quanti gradi dovrebbe essere ciascun settore circolare in un diagramma a torta.

b) Diagramma a torta, prima "a mano" sul quaderno poi con un foglio elettronico

2) Percentuale approssimata, in peso, degli elementi chimici sulla crosta terrestre (dati presi da http://hyperphysics.phy-astr.gsu.edu e provenienti da *Lutgens and Tarbuck, Essentials of Geology*)

Element	*Approximate % by weight*
Oxygen	46,6
Silicon	27,7
Aluminum	8,1
Iron	5,0
Calcium	3,6
Sodium	2,8
Potassium	2,6
Magnesium	2,1
All others	1,5

a) Determina l'angolo al centro corrispondente a ciascun settore
b) poi traccia sul tuo quaderno il diagramma a torta
c) e infine costruisci lo stesso diagramma con un foglio elettronico, stampa, appiccicalo accanto al tuo.

3) Lo stesso sito http://hyperphysics.phy-astr.gsu.edu riporta anche la seguente tabella, tratta da *Biology, Life on Earth* di Teresa e Gerald Audesirk:

Element	*Symbol*	*Atomic Number*	*Percent in Universe*	*Percent in Earth*	*Percent in Human Body*
Hydrogen	H	1	91	0,14	9,5
Helium	He	2	9	Trace	Trace
Carbon	C	6	0,02	0,03	18,5
Nitrogen	N	7	0,04	Trace	3,3
Oxygen	O	8	0,06	47	65
Sodium	Na	11	Trace	2,8	0,2
Magnesium	Mg	12	Trace	2,1	0,1
Phosphorus	P	15	Trace	0,07	1
Sulfur	S	16	Trace	0,03	0,3
Chlorine	Cl	17	Trace	0,01	0,2
Potassium	K	19	Trace	2,6	0,4
Calcium	Ca	20	Trace	3,6	1,5
Iron	Fe	26	Trace	5	Trace

Utilizza l'ultima colonna per tracciare un diagramma a torta sulla composizione chimica del corpo umano.

4) Dati da US Geological - I maggiori produttori d'oro nel 2007:

Nazione	*Tonnellate*
Australia	280
Sud Africa	270
Cina	250
Stati Uniti	240
Peru	170
Russia	160
Indonesia	120
Canada	100

Istogramma (=diagramma a barre verticali), prima sul quaderno e poi al computer con un foglio elettronico.

5) Da http://nineplanets.org, alcuni dati sui pianeti del sistema solare.

Planet	Distance from Sun (1000 km)	Radius (Km)	Mass (Kg)	Discoverer-Date
Mercury	57.910	2 439	3,30e23 (NOTA)	
Venus	108.200	6 052	4,87e24	
Earth	149.600	6 378	5,98e24	
Mars	227.940	3 397	6,42e23	
Jupiter	778.330	71 492	1,90e27	
Saturn	1.426.940	60 268	5,69e26	
Uranus	2.870.990	25 559	8,69e25	Herschel - 1781
Neptune	4.497.070	24 764	1,02e26	Galle - 1846
Pluto	5.913.520	1 160	1,31e22	Tombaugh - 1930

Realizza un diagramma a barre orizzontali, con un foglio elettronico, per la colonna del raggio.

NOTA: 3,30e23 significa (notazione esponenziale): $3,30 \cdot 10^{23}$

6) Mortalità infantile in Italia:
numero morti nel primo anno di vita per ogni 1000 nati vivi.

Traccia la "serie storica".

1865	30,3
1905	27,3
1935	21,6
1965	16,6
1985	13,9
2002	4,1

7) Una volante della polizia si apposta per rilevare la velocità delle macchine che transitano.
Gli esiti di una serie di 60 controlli sono i seguenti (in Km/h):

57	61	52	48	68	48	55	56	72	49	50	55
54	61	58	63	64	61	47	52	53	59	48	54
56	60	51	51	50	78	67	61	58	55	59	53
62	54	49	45	56	60	51	52	60	54	55	51
48	57	56	55	58	53	59	70	74	64	81	52

Suddividi i dati in classi di frequenza, traccia l'istogramma.

8) I punteggi
a) base b) dell'esame scritto c) dell'esame orale
alla prova finale, per una classe di 18 candidati, sono stati i seguenti.
Con un foglio elettronico, costruisci la rappresentazione più opportuna.

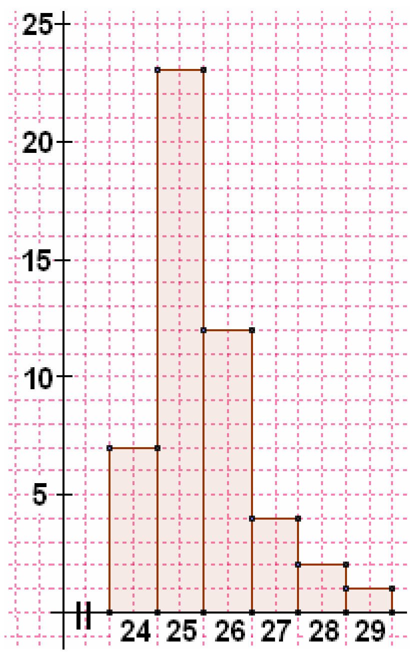
a)	15	18	24	25	21	17	19	16	20	21	23	20	16	18	18	22	25	24
b)	29	32	33	32	30	29	28	28	34	39	42	37	32	32	33	35	45	43
c)	20	22	24	30	22	24	27	19	25	24	24	20	18	19	21	25	30	30

9) Un gruppo di ingegneri ha dichiarato
di aver conseguito la laurea specialistica
alle età che sono indicate dall'istogramma qui a fianco.

a) Quante persone sono state intervistate?

b) Quali sono la frequenza assoluta, relativa e percentuale
di coloro che si sono laureati non prima dei 27 anni?

c) Qual è, sul totale, la percentuale di ingegneri
che si è laureata prima dei 26 anni?

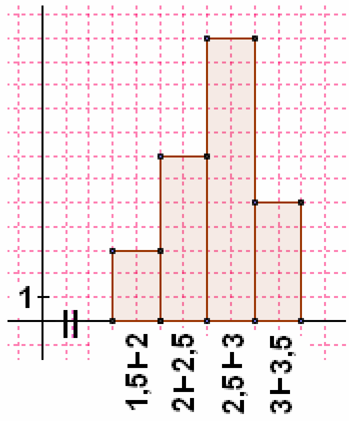


10) Un insegnante di Educazione Fisica fa effettuare le prove preliminari ai suoi giovani allievi, e registra nell'istogramma qui riportato le misure raggiunte nel salto in lungo.

Si domanda la frequenza
- ❑ assoluta
- ❑ relativa (arrotondata ai centesimi)
- ❑ e percentuale (arrotondata alle unità)

di coloro che hanno saltato almeno 3 metri.

La regola per arrotondare è richiamata alla pag. 66 di questo volume.



11) Rappresenta graficamente in modo adeguato alcuni degli esiti 1) … 15) del "questionario del curiosone" di pag. 8

12) La tabella qui a fianco riporta:
in 2^a colonna, i nati in totale (valore medio annuo in migliaia);
in 3^a colonna, i nati per 1000 abitanti
(i dati provengono da www.istitutodeglinnocenti.it).

a) Con un foglio elettronico, traccia la serie storica

b) Quante nascite in totale si sono avute, pressappoco, in Italia, nell'arco di tempo dal 1931 al 1940?

c) Come si concilia il fatto che nella seconda colonna la differenza relativa dei dati non sia forte mentre in terza colonna decisamente sì?

1861-70	947	37,6
1891-900	1099	35,0
1911-20	972	27,2
1931-40	1008	23,6
1951-60	872	17,9
1961-70	953	18,3
1971-80	791	14,2
1981-85	600	10,6
2005	554	9,5

13) 4 associazioni A, B, C, D di appassionati hanno contribuito a una mostra di minerali rispettivamente con 45, 72, 18 e 29 pezzi. Traccia sul quaderno, e trova poi il modo di realizzare pure con un foglio elettronico, un diagramma formato da un'unica striscia orizzontale ⊢⊢⊢⊣⊣, che visualizzi la ripartizione.

14) Da ISTAT, "Italia in cifre":

DISTRIBUZIONE DELLA POPOLAZIONE DI 6 ANNI E PIÙ PER TITOLO DI STUDIO Censimenti 1951-2001, composizioni percentuali

	Laureati	Diplomati	Con licenza media	Con licenza elementare	Alfabeti privi di titoli di studio	Analfabeti
1951	1,0	3,3	5,9	30,6	46,3	12,9
1961	1,3	4,3	9,6	42,3	34,2	8,3
1971	1,8	6,9	14,7	44,3	27,1	5,2
1981	2,8	11,5	23,8	40,6	18,2	3,1
1991	4,2	18,2	30,7	32,6	12,2	2,1
2001	7,1	26,2	30,1	25,4	9,7	1,5

Con un foglio elettronico, traccia diversi tipi di diagramma per illustrare le situazioni e confrontarle fra loro. Esempi:

a) un diagramma a barre verticali che presenti il livello di istruzione nel 2001
b) un diagramma a barre orizzontali che faccia vedere le percentuali dei diplomati nei vari anni
c) un diagramma a torta che evidenzi il livello di istruzione nel 1951
d) un diagramma cartesiano (serie storica) che mostri, in simultanea, l'evoluzione negli anni del numero di laureati e di analfabeti

15) Richieste analoghe a quelle dell'esercizio precedente per la tabella che segue (dati ISTAT).

MEZZI DI TRASPORTO UTILIZZATI PER ANDARE AL LAVORO
Anni 1997-2008, per 100 occupati di 15 anni e più che escono di casa per andare al lavoro

	1997	1999	2001	2003	2005	2007	2008
Treno	2,1	2,4	2,6	2,3	2,6	2,9	2,9
Tram, bus	4,9	5,5	5,0	5,0	5,0	5,5	4,9
Metropolitana	1,8	1,9	1,8	2,2	2,2	2,5	2,5
Pullman	2,0	2,4	1,9	2,1	3,0	3,2	2,9
Automobile	72,0	72,0	75,0	75,2	76,8	74,4	75,7
Moto, ciclomotore	4,0	4,7	4,3	5,0	4,2	4,4	4,6
Bicicletta	2,6	3,6	2,7	3,0	2,9	3,2	3,1
A piedi	13,6	11,8	11,6	10,9	11,1	11,2	11,1

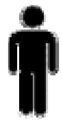
16) Nelle due filiali che una banca ha recentemente aperto in una cittadina, l'ispettore inviato dalla direzione centrale organizza un sondaggio sul grado di soddisfazione dei clienti riguardo ai vari servizi.
Viene raccolto il parere di 47 persone nella prima filiale e di 35 nella seconda.

Filiale A	Poco soddisfatto	Sufficientemente soddisfatto	Molto soddisfatto
Cortesia ed efficienza degli impiegati	2	31	14
Qualità della consulenza finanziaria	9	34	4
Tempi di attesa allo sportello	2	23	22

Filiale B	Poco soddisfatto	Sufficientemente soddisfatto	Molto soddisfatto
Cortesia ed efficienza degli impiegati	5	28	2
Qualità della consulenza finanziaria	8	23	4
Tempi di attesa allo sportello	5	27	3

Quale rappresentazione grafica ti sembra più adeguata a illustrare visivamente la situazione ai fini di un'analisi e di un confronto? Realizzala al computer con un foglio elettronico.

17) Considera la tabella sottostante (ISTAT , "Italia in cifre", anno 2007) e traduci un suo aspetto a tua scelta (ad esempio, potresti prendere i totali per settore sull'intera Italia …) in un ideogramma basato sull'icona qui a fianco (puoi scegliere se restringere/allargare l'icona, oppure in alternativa riportarla più volte).



OCCUPATI PER POSIZIONE PROFESSIONALE, SETTORE DI ATTIVITÀ E AREA GEOGRAFICA
Anno 2008, migliaia di persone

	Nord	Centro	Mezzogiorno	Italia
AGRICOLTURA				
Dipendenti	114	49	263	425
Indipendenti	232	67	172	470
Totale	**346**	**115**	**434**	**895**
INDUSTRIA				
Dipendenti	3.340	984	1.175	5.499
Indipendenti	818	309	329	1.456
Totale	**4.157**	**1.293**	**1.504**	**6.955**
SERVIZI				
Dipendenti	5.627	2.568	3.326	11.522
Indipendenti	1.936	880	1.217	4.033
Totale	**7.563**	**3.448**	**4.543**	**15.555**
TOTALE GENERALE				
Dipendenti	9.081	3.601	4.764	17.446
Indipendenti	2.985	1.256	1.718	5.959
Totale	**12.066**	**4.857**	**6.482**	**23.405**

18) I seguenti dati provengono dalla WAN, World Association of Newspapers: numero di copie di quotidiani diffuse giornalmente ogni 1000 abitanti nell'anno 2002.

a) Vuoi scegliere 5 o 6 nazioni nel lungo elenco e tracciare un diagramma a barre?

b) E che ne diresti di un ideogramma a icona singola (rimpicciolita o ingrandita)? O di un ideogramma con icone ripetute (doppio numero di icone = doppia diffusione del quotidiano)?

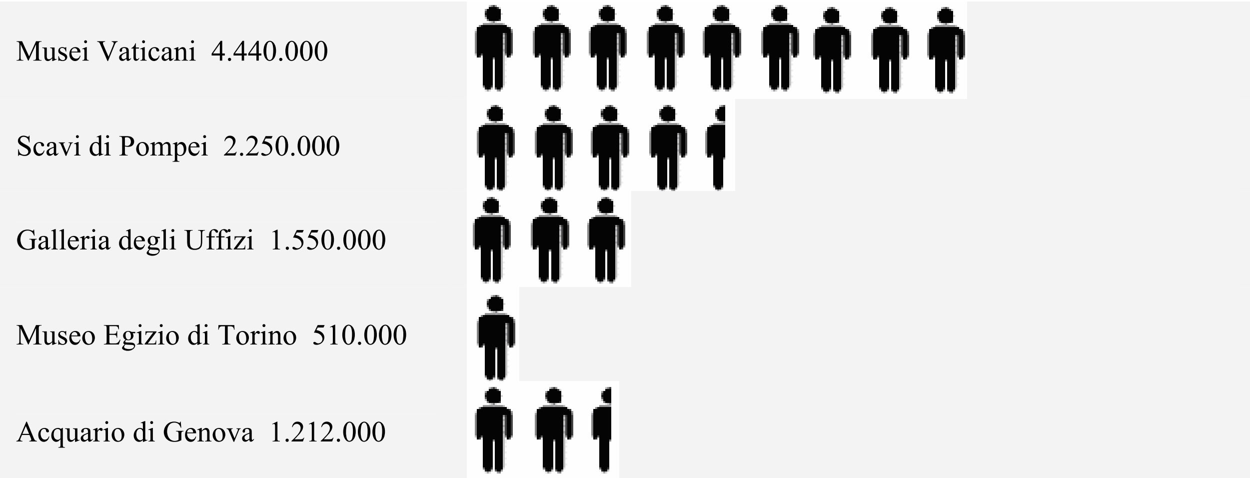
1. Norvegia	705	14. Usa	274	27. Francia	181	40. Cipro	94
2. Giappone	664	15. Nuova Zelanda	259	28. Belgio	175	41. Polonia	92
3. Finlandia	544	16. Estonia	234	29. Russia	146	42. Portogallo	91
4. Svezia	543	17. Thailandia	234	30. Turchia	131	43. Grecia	81
5. Svizzera	444	18. Irlanda	233	31. Cina	130	44. Brasile	64
6. Islanda	393	19. Slovenia	214	32. Croazia	128	45. Argentina	56
7. Regno Unito	383	20. Malesia	209	33. Italia	128	46. India	48
8. Germania	371	21. Rep. Ceca	206	34. Spagna	120	47. Sud Africa	40
9. Danimarca	371	22. Bulgaria	203	35. Costarica	120	48. Sri Lanka	35
10. Austria	363	23. Australia	202	36. Rep. Slovacca	117	49. Indonesia	31
11. Olanda	363	24. Ungheria	199	37. Ucraina	105	50. Kenya	14
12. Lussemburgo	339	25. Canada	189	38. Filippine	99		
13. Singapore	331	26. Lettonia	184	39. Libano	96		

19) La più recente serata del sabato sera (intervista-flash ai ragazzi in uscita da una scuola superiore).

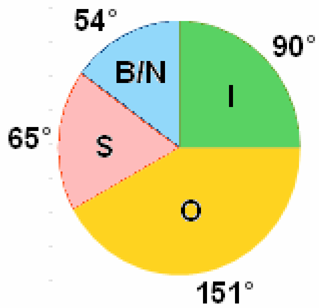
Discoteca	*Birreria*	*Cinema*	*Casa di amici*	*Altro*
43	98	26	57	75

Rappresenta gli esiti del sondaggio
a) con un diagramma a barre
b) con un ideogramma a figure ripetute (magari utilizzando come icona l'omino dell'esercizio 17)
c) con un diagramma costituito da un'unica striscia orizzontale

20) La seguente rappresentazione per ideogrammi è relativa al numero (approssimativo) di visitatori totali di alcuni musei italiani, nell'anno 2008 (dati dal sito del Touring Club).



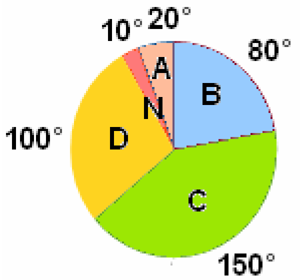
Guardando solo l'ideogramma e non i numeri, stabilisci quale è stata all'incirca la percentuale delle visite all'Acquario di Genova, rispetto alla Galleria degli Uffizi.



21) Determina le percentuali approssimative di voti totalizzate dai tre partiti →
- ❑ *Insieme per il buon governo* (angolo di 90°)
- ❑ *Onestà e competenza* (approssimativamente 151°)
- ❑ *Siamo con voi* (all'incirca 65°)

(ci sono state anche molte schede *Bianche o Nulle* in questa votazione)

22) Se in totale i ragazzi della scuola sono 324, e il diagramma qui a fianco ne riporta la ripartizione fra coloro che, promossi, hanno avuto quest'anno in pagella finale la media dei voti compresa fra 6 e 7 (escluso), fra 7 e 8 (escluso), fra 8 e 9 (escluso), non inferiore a 9, oppure non sono stati ammessi alla classe successiva, determinare il numero dei ripetenti e degli eccellenti.



N: Non ammessi
D: $6 \le m < 7$
C: $7 \le m < 8$
B: $8 \le m < 9$
A: $m \ge 9$

23) Un *bed and breakfast* di una località montana ha suddiviso la sua attività in tre periodi: da marzo a giugno; luglio+agosto; da settembre a novembre.
Per l'anno passato, sono andati persi i dati di ciascun giorno ma rimangono i dati complessivi dei 3 periodi: 598 ospiti da marzo a giugno; 895 in luglio+agosto; 327 da settembre a novembre.
Costruisci un istogramma che illustri la situazione, con
- ❑ tre intervalli in ascissa proporzionali alle rispettive durate
- ❑ *aree* dei rettangoli proporzionali al numero di persone ospitate.

24) Un'azienda, in occasione del ventennale della sua fondazione, stila un quadro del numero di dipendenti ripartendoli per anzianità di servizio. Vengono distinte 3 categorie:

fino a 2 anni	25
più di 2 e fino a 5	45
più di 5 e fino a 20	48

Disegna un istogramma, nel quale siano le *aree* a rappresentare il numero di dipendenti in una determinata fascia di anzianità.

25) Illustra la relazione esistente nella tua classe fra numero di scarpe e altezza in centimetri compilando, tramite un foglio elettronico, una *tabella a doppia entrata* che porti, per i vari numeri di scarpe e opportune fasce di altezze, la frequenza, nella tua classe, di quel doppio dato.

26) Nel 2010, secondo il CDIAC, Carbon Dioxide Information Analysis Center (http://cdiac.ornl.gov/), il tristissimo primato **dell'emissione di anidride carbonica nell'aria** spettava ai 10 paesi seguenti, per i quali viene specificato anche il numero di tonnellate che si stima abbiano diffuso in quell'anno:

Cina	USA	India	Russia	Giappone	Germania	Iran	Corea	Canada	Arabia
8240958	5492170	2069738	1688688	1138432	762543	574667	563126	518475	493726

Poiché questi dati non comprendono gli altri paesi del mondo, che pure contribuiscono all'inquinamento globale, qui un diagramma a barre è preferibile rispetto a un diagramma a torta. Realizzalo al computer.

… Una delle cose che possiamo fare, per combattere contro questa folle corsa del mondo all'autodistruzione, è abituarci al pensiero che una vita semplice, con una forte autoriduzione dei consumi, è la sola compatibile con la finitezza delle risorse della Terra, ed è anche di gran lunga più degna e gioiosa.

MDF PIL
MOVIMENTO PER LA DECRESCITA FELICE
http://decrescitafelice.it/

27) La causa primaria della deforestazione dell'Amazzonia sta nell'abbattimento di vasti tratti di foresta che viene rimpiazzata con coltivazioni destinate al nutrimento degli animali fornitori di **carne**.

CAUSA DI DEFORESTAZIONE	PERCENTUALE
Allevamento di animali e coltivazioni relative	**60-70%**
Agricoltura di sussistenza e su piccola scala	30-40%
Agricoltura commerciale su vasta scala	1-2%
Taglio di alberi per legname, legale e illegale	1-2%
Incendi, miniere, strade, dighe, urbanizzazione	2-4%

La tabella qui a fianco è tratta da Mongabay.com, Rhett A. Butler - San Francisco, CA., 2000-2007

Si stima che nel 1977, pur essendo già iniziata l'opera di decimazione, rimanessero circa 3 955 870 km^2 di foresta in Amazzonia. La tabella indica i km^2 approssimativamente persi negli anni successivi.

1978-1987	21130	1992	13786	1997	13227	2002	21394
1988	21050	1993	14896	1998	17383	2003	25247
1989	17770	1994	14896	1999	17259	2004	27423
1990	13730	1995	29059	2000	18226	2005	18846
1991	11030	1996	18161	2001	18165	2006	14109

Con un foglio elettronico, traccia:
- □ un diagramma a torta per le cause della deforestazione amazzonica;
- □ una serie storica che rappresenti l'estensione della foresta nel periodo 1977-2006.

Il problema delle risorse che vanno in fumo a causa degli allevamenti intensivi è ENORME.

Se le immense estensioni di terreno utilizzate per dar da mangiare agli "animali da carne" fossero convertite in coltivazioni di ottima verdura o cereali da destinare all'alimentazione umana, TANTI PROBLEMI DI DEFORESTAZIONE E DEGRADO AMBIENTALE SAREBBERO RISOLTI: basti pensare che un ettaro coltivato a patate o a riso è in grado di provvedere al nutrimento annuo rispettivamente di 22 e 19 persone, mentre quando lo stesso ettaro è destinato alla produzione di vegetali per l'ingrasso dei manzi, la carne che se ne ricava può bastare per UNA SOLA persona.

Leggi a proposito questo interessantissimo articolo, che analizza tutti gli aspetti della questione: ⇨

E ancor prima … ma che diritto ha la arrogante e presuntuosa razza umana, di sfruttare senza pietà, imprigionare, umiliare e trattare alla stregua di oggetti degli esseri senzienti capaci di provare dolore e terrore e impossibilitati a difendersi???

Una dieta vegana o pesco-vegana (*) oltretutto è IDEALE PER LA SALUTE! Vedi ad esempio il riquadro qui a sinistra.

(*) La dieta pesco-vegana consente il consumo di pesci purché NON di allevamento

Pitagora, Seneca, Leonardo, Gandhi, Tolstoj, Einstein, Paul McCartney … … Ragione, pietà e spirito nelle parole dei grandi vegetariani-vegani ⇨

8. GLI INDICI DI POSIZIONE (o “di centralità”)

A) LE MEDIE “FERME”

(una media si dice “**ferma**” se **il suo valore varia senz’altro, qualora uno solo dei termini in gioco cambi**).

- HO COMINCIATO AD ALLENARMI PER LA MARATONA DI NEW YORK!

  Lunedì ho corso per 4 km, martedì per 6,5 km, mercoledì per 8 km, giovedì per 2,5 km, venerdì per 5 km, sabato per 7,5 km, domenica per 8,5 km.
  Quanti km ho percorso in media al giorno?

  Una “**media**” **fra più numeri**, che esprimono quantità della stessa specie, **è un numero avente la proprietà di essere compreso (“medio” = “che sta in mezzo”) tra il minore e il maggiore dei numeri dati**.

  La risposta alla nostra domanda NON sarà però, evidentemente,
  uno qualsiasi fra i valori compresi tra 2,5 e 8,5
  e nemmeno *il numero esattamente intermedio* fra 2,5 e 8,5 (che sarebbe 5,5)!

  Ragioniamo. Noi vogliamo trovare quel numero x tale che,
  se in ognuno dei 7 giorni della settimana io avessi corso ogni volta per esattamente x km,
  la distanza complessiva percorsa in tutta la settimana sarebbe stata la medesima! Allora

  $$7x = 4+6{,}5+8+2{,}5+5+7{,}5+8{,}5$$

  da cui

  $$x = media\ dei\ km = n°\ di\ km\ \ percorsi\ mediamente\ in\ 1\ giorno = \frac{4+6{,}5+8+2{,}5+5+7{,}5+8{,}5}{7} = \frac{42}{7} = 6$$

  In effetti, se ogni giorno della settimana il mio percorso fosse stato di esattamente 6 km,
  complessivamente nella settimana mi sarei allenato per un totale di $6+6+6+6+6+6+6 = 42$ km,
  esattamente come è $4+6{,}5+8+2{,}5+5+7{,}5+8{,}5 = 42$.

- E’ ragionevole supporre che la prestazione complessiva di uno studente, che ha preso diversi voti

  $$v_1, v_2, \ldots, v_n,$$

  possa essere bene rappresentata dal particolare voto v che, se fosse stato preso, *sempre quello*, tutte le n volte, avrebbe dato luogo alla *stessa somma di voti*. Dunque

  $$nv = v_1 + v_2 + \ldots + v_n$$

  $$v = media\ dei\ voti = \frac{v_1 + v_2 + \ldots + v_n}{n}$$

  Ad es., se quello studente ha preso i 5 voti $6\ \ 7\ \ 7\ \ 8\ \ 6$, la sua media è stata $\frac{6+7+7+8+6}{5} = \frac{34}{5} = 6{,}8$.

  E in effetti, se quello studente avesse invece preso come voti successivi $6{,}8\quad 6{,}8\quad 6{,}8\quad 6{,}8\quad 6{,}8$
  la somma dei suoi voti, ossia la sua “prestazione totale”, sarebbe stata $6{,}8+6{,}8+6{,}8+6{,}8+6{,}8 = 34$
  esattamente uguale alla prestazione effettiva totale $6+7+7+8+6 = 34$

  Si dice **MEDIA ARITMETICA** M fra n numeri $x_1, x_2, \ldots, x_n$, il numero

  $$\boxed{\boldsymbol{M = media\ aritmetica = \frac{x_1 + x_2 + \ldots + x_n}{n}}}\ .$$

  La media aritmetica fra più valori, è uguale alla loro somma, divisa per il numero dei valori stessi; ed è quel nuovo valore il quale, se sostituito al posto di ciascuno dei singoli valori in gioco, ne lascerebbe invariata la somma.

ESERCIZI

1) Verifica che se ho mangiato il minestrone 3 volte in novembre, 8 in dicembre, 8 in gennaio, 5 in febbraio, e 9 volte in totale gli altri mesi dell’anno, in media è come se l’avessi mangiato 2,75 volte al mese.

2) Calcola la media del numero di scarpe, fra i compagni di classe: a) maschi; b) femmine; c) tutti.
   In generale la media c) *non* coincide con la media delle due medie a) e b) … a meno che …

3) Con riferimento alle classi I A, I B di cui a pagina 354, con un foglio elettronico rappresenta la “serie storica” delle medie aritmetiche dei punteggi (mettendo, ogni anno, assieme le due classi a formare un unico gruppo).
   Le “lagnanze” di cui si parla nella stessa pagina sono giustificate, a giudicare da questa successione di medie?

La “media” di cui ci siamo occupati fin qui è stata la media “aritmetica”
(anche se sovente, per brevità, l’aggettivo viene lasciato sottinteso);
in effetti, **ci sono altri tipi di “medie”**, oltre a questa, e ora andremo brevemente a illustrarli.

- Se il COSTO DI UNA MATERIA PRIMA è aumentato:
  - del 5% nel 2001 (s'intende: da inizio a fine anno);
  - del 6% nel 2002;
  - dell'8% nel 2003;
  - dell'8% ancora nel 2004;
  - e del 4% nel 2005,

  a) di quanto è aumentato *complessivamente* nel quinquennio 2001-2005?

  b) E di quanto è aumentato *mediamente* ogni anno, in questo quinquennio?

Ragioniamo.

a) Se il prezzo all'inizio del 2001 era 100,

- alla fine del 2001 è diventato $100\cdot\left(1+\frac{5}{100}\right)=100\boxed{\cdot 1{,}05}=105$
- alla fine del 2002 è diventato $105\cdot\left(1+\frac{6}{100}\right)=105\boxed{\cdot 1{,}06}=111{,}3$
- alla fine del 2003 è diventato $111{,}3\cdot\left(1+\frac{8}{100}\right)=111{,}3\boxed{\cdot 1{,}08}=120{,}2$ (*circa*)
- alla fine del 2004 è diventato *circa* $120{,}2\cdot\left(1+\frac{8}{100}\right)=120{,}2\boxed{\cdot 1{,}08}=$ *circa* $129{,}8$
- alla fine del 2005 è diventato *circa* $129{,}8\cdot\left(1+\frac{4}{100}\right)=129{,}8\boxed{\cdot 1{,}04}=$ *circa* 135;

è perciò aumentato, questo prezzo, da inizio 2001 a fine 2005, complessivamente intorno al 35%.

b) E *mediamente*, quanto è aumentato?
Noi *cerchiamo* in questo momento *una percentuale annua x tale che,*
se *l'aumento fosse stato ogni anno esattamente dell'x%, si sarebbe raggiunto il medesimo prezzo finale.*
Quindi

$$\cancel{100}\cdot\left(1+\frac{x}{100}\right)\left(1+\frac{x}{100}\right)\left(1+\frac{x}{100}\right)\left(1+\frac{x}{100}\right)\left(1+\frac{x}{100}\right)=\cancel{100}\cdot\left(1+\frac{5}{100}\right)\left(1+\frac{6}{100}\right)\left(1+\frac{8}{100}\right)\left(1+\frac{8}{100}\right)\left(1+\frac{4}{100}\right)$$

$$\left(1+\frac{x}{100}\right)^5=1{,}05\cdot 1{,}06\cdot 1{,}08\cdot 1{,}08\cdot 1{,}04$$

$$1+\frac{x}{100}=\sqrt[5]{1{,}05\cdot 1{,}06\cdot 1{,}08\cdot 1{,}08\cdot 1{,}04}$$

$$1+p=\sqrt[5]{1{,}05\cdot 1{,}06\cdot 1{,}08\cdot 1{,}08\cdot 1{,}04}\approx\sqrt[5]{1{,}35}\approx 1{,}062 \quad \textit{da cui} \quad p=\sqrt[5]{1{,}05\cdot 1{,}06\cdot 1{,}08\cdot 1{,}08\cdot 1{,}04}-1\approx 0{,}062=6{,}2\%$$

Vuol dire che, se quel prezzo iniziale fosse aumentato ogni anno del $6{,}2\%$, si sarebbe raggiunto,
dopo un quinquennio, lo stesso prezzo finale che si è ottenuto con gli aumenti del 5%, 6%, 8%, 8%, 4%.
Vuoi provare a verificarlo col calcolo?

Presi dunque i valori $1{,}05$ $1{,}06$ $1{,}08$ $1{,}08$ $1{,}04$
che davano il numero per cui moltiplicare il prezzo all'inizio dell'anno, onde ottenere il prezzo alla fine,
la "media sul quinquennio" di questi moltiplicatori è la radice quinta del loro prodotto
(e non, come nel caso della media aritmetica, la quinta parte della loro somma)!

Si definisce "MEDIA GEOMETRICA" fra più valori, quel nuovo valore il quale,
se sostituito al posto di ciascuno dei singoli valori in gioco, ne lascerebbe invariato il prodotto.
Si dimostra facilmente che la media geometrica fra *n* valori $x_1, x_2, \dots, x_n$ è data da

$$\boldsymbol{M_G = \textit{media geometrica} = \sqrt[n]{x_1\cdot x_2\cdot\ldots\cdot x_n}}$$

Infatti, poiché si desidera che questa media, sostituita a ciascuno dei valori, non ne alteri il prodotto:

$$x_1\cdot x_2\cdot\ldots\cdot x_n=M_G\cdot M_G\cdot\ldots\cdot M_G$$
$$M_G{}^n=x_1\cdot x_2\cdot\ldots\cdot x_n$$
$$M_G=\sqrt[n]{x_1\cdot x_2\cdot\ldots\cdot x_n}$$

Come indicazione generale, possiamo dire che la media geometrica
si utilizza quando i dati sono tali che per essi l' "operazione regina"
è il *prodotto*, piuttosto che la somma.
Quindi, in un contesto di tassi di interesse bancari,
o di aumento o diminuzione (rara ...) dei prezzi,
o di incremento o decremento del PIL, o di tasso di crescita di una popolazione,
dobbiamo aspettarci di incontrare medie geometriche piuttosto che aritmetiche.

Robert Kennedy *Discorso sul PIL* *(marzo 1968)*

- Facciamo UN ALTRO ESEMPIO DI NATURA DIVERSA, molto significativo.

$v=\frac{s}{t}$
$s=vt$
$t=\frac{s}{v}$

Se ho percorso in auto un totale di 100 km, la prima metà andando ai 100 km/h
e la metà successiva, dopo aver visto un brutto incidente, agli 80 km/h soltanto,
quale è stata la mia velocità media?

Rispondere che è stata la media aritmetica delle due velocità,
quindi $\frac{100+80}{2}=90$ km/h, sarebbe magari istintivo … ma clamorosamente SBAGLIATO.

Infatti è logico partire dal presupposto che per "velocità media", in questo contesto, si debba intendere *quella velocità la quale, se mantenuta costante per tutto il tragitto di 100 km, mi avrebbe permesso di coprirlo nel medesimo tempo.*

E *quanto tempo* ci ho messo a fare i miei 100 km,
andando per 50 km ai 100 all'ora e per 50 km agli 80 all'ora?

Vediamo.
La prima metà del percorso ha richiesto un tempo, in ore, uguale a

$$\frac{s}{v}=\frac{50}{100}=\frac{1}{2}=0,5 \text{ (mezz'ora, dunque),}$$

mentre la seconda metà ha richiesto un numero di ore dato da

$$\frac{s}{v}=\frac{50}{80}=0,625 \text{ (0,625 ore, o anche: 37 minuti e mezzo).}$$

Il tempo totale per coprire il tragitto di 100 km è stato perciò di ore $0,5+0,625=1{,}125\ (1h\ 7'\ 30'')$.

Ma se una distanza di 100 km venisse percorsa ad andatura costante in ore 1,125
vorrebbe dire che quella velocità costante è di

$$\text{km/h } \frac{100}{1{,}125} \approx \text{km/h } 88,9$$

Quindi in questo caso per calcolare la "velocità media"
NON si deve fare la "media aritmetica delle due velocità"!

Si deve invece procedere
1) direttamente col ragionamento e col calcolo, come abbiamo fatto noi;
2) oppure (lo si potrebbe dimostrare) calcolando la cosiddetta *media armonica* delle velocità.

$$\boxed{\boldsymbol{M_A = \text{media armonica} = \frac{1}{\dfrac{\dfrac{1}{x_1}+\dfrac{1}{x_2}+\ldots+\dfrac{1}{x_n}}{n}}}}$$

Si definisce "MEDIA ARMONICA" fra *n* valori, quel nuovo valore il quale, se sostituito al posto di ciascuno dei singoli valori in gioco, ne lascerebbe invariata la somma dei reciproci. Essa coincide col reciproco della media aritmetica dei reciproci dei valori in gioco.

Si può far vedere (vuoi provarci?) che, se una data distanza d viene suddivisa
in n tratti tutti uguali fra loro (ogni tratto ha quindi lunghezza d/n),
e questi tratti vengono percorsi alle velocità $v_1, v_2, \ldots, v_n$ rispettivamente,
per cui il viaggio richiede un certo tempo totale t,
allora la "velocità media", intesa come la velocità costante
alla quale occorrerebbe muoversi per percorrere la stessa distanza d nello stesso tempo t,

- *non dipende dalla distanza* d
- *ed è data dalla media armonica delle velocità:*

$$\boxed{v=\frac{1}{\dfrac{\dfrac{1}{v_1}+\dfrac{1}{v_2}+\ldots+\dfrac{1}{v_n}}{n}}}$$

OSSERVAZIONE
Invece, se noi avessimo un tempo di viaggio fissato t suddiviso in n intervalli di ugual durata t/n,
e in questi n intervalli uguali di tempo si procedesse alle velocità costanti $v_1, v_2, \ldots, v_n$,
percorrendo una determinata distanza totale d,
la velocità costante alla quale procedere se si desidera, sempre nel tempo t, percorrere la stessa distanza d
non dipenderebbe da t e sarebbe data dalla *media aritmetica* delle velocità $v'=\frac{v_1+v_2+\ldots+v_n}{n}$ (dimostralo!)

- Ancora:

Si definisce "MEDIA QUADRATICA" fra più valori, quel nuovo valore il quale,
se sostituito al posto di ciascuno dei singoli valori in gioco,
ne lascerebbe invariata la somma dei quadrati.

Si dimostra che la media quadratica fra n valori $x_1, x_2, \dots, x_n$ è data da

$$\boxed{M_Q = \text{media quadratica} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}}$$

Cercando di trarre le conclusioni da questo tentativo di GENERALIZZAZIONE DEL CONCETTO DI "MEDIA", potremo dire (traducendo in forma più semplice una definizione di Oscar Chisini, 1889-1967) che
se si hanno n valori $x_1, x_2, \dots, x_n$ di una grandezza, si può parlare di "media" ogniqualvolta si desidera determinare un valore x che, qualora venisse sostituito al posto di ciascuno dei valori dati, ne lascerebbe invariata, a seconda del tipo di "media":

- **la somma;**
- **o il prodotto;**
- **…**
- **oppure una qualunque determinata loro "funzione",**
  ossia grandezza che dipenda, secondo una legge ben definita, dalle grandezze date.

UNA MEDIA
a) … IN PARTE DISTRUGGE, E IN PARTE RIESCE A MANTENERE L'INFORMAZIONE;
b) … E' UN VALORE "TEORICO";
c) … E CI DA' SOLO QUELLO CHE DA LEI SAPPIAMO DI POTERCI ASPETTARE!

a) Una media, di qualsiasi tipo essa sia, cerca di sintetizzare in un singolo numero un'informazione relativa a una pluralità di dati (sovente, a *tantissimi* dati).
Evidentemente, essa *non* può pretendere di condensare in sé tutto il contenuto informativo insito nell'insieme *effettivo* dei dati;
passando alla "media" tale contenuto in gran parte va perso …
e tuttavia qualcosa, peraltro di molto importante, rimane.

b) Una media è un valore "TEORICO", nel senso che ben raramente coincide con uno dei dati in questione (e, se anche ciò avviene, questo fatto non è comunque particolarmente interessante).

c) Una media "CI DA' SOLO QUELLO CHE DA LEI CI ASPETTIAMO",
nel senso che, ad esempio,

- una media *aritmetica* ci dà quel valore che, se venisse sostituito a ciascuno dei dati, ne lascerebbe inalterata la somma;
- una media *geometrica* ci dà quel valore che, se venisse sostituito a ciascuno dei dati, ne lascerebbe inalterato il prodotto;
- una media *armonica* ci dà quel valore che, se venisse sostituito a ciascuno dei dati, ne lascerebbe inalterata la somma dei reciproci;
- eccetera.

LA PIENA COMPRENSIONE DEL SIGNIFICATO DI UNA "MEDIA",
cioè del tipo di informazione che essa ci dà,
È LEGATA ALLA CONSAPEVOLEZZA DI
"QUAL È LA QUANTITÀ CHE RESTEREBBE INALTERATA
SE AL POSTO DI CIASCUNO DEI DATI
SI SOSTITUISSE LA MEDIA DEI DATI STESSI"

Si dice che la media in esame "CONSERVA"
quella determinata quantità:
ad esempio, la media aritmetica "conserva la somma",
perché, se venisse sostituita al posto di ciascuno dei dati,
la somma di questi non muterebbe.

- La media con cui lo studente ha quasi sempre a che fare è la media *aritmetica* (quando si dice semplicemente "*media*", *senza aggettivi*, è alla *media aritmetica* che ci si riferisce).
- Per la precisione, nelle pagine precedenti, avremmo dovuto scrivere "media aritmetica *semplice*", "media geometrica *semplice*", "media armonica *semplice*", … per distinguere le medie introdotte dalle corrispondenti medie "*ponderate*".
  Alla media aritmetica "ponderata" faremo cenno fra breve.

In un'indagine statistica, o in un diagramma statistico, i "dati" di cui fare la "media" sono le "modalità" (è ovvio che ha senso farne la media soltanto se queste sono espresse numericamente); ciascuna modalità viene contata tante volte quant'è la sua frequenza nella popolazione statistica in esame.

Ad esempio, nella rilevazione del numero di figli da 0 a 10 anni di un gruppo di 20 famiglie, la distribuzione di frequenze potrebbe essere

0	1	2	3
7	8	4	1

E in questo caso avrebbe senso fare la media aritmetica del numero dei figli, che sarebbe

$$M=\frac{0+0+0+0+0+0+0+1+1+1+1+1+1+1+1+2+2+2+2+3}{20}=\frac{19}{20}=0,95$$

Se i dati sono stati ripartiti in intervalli ossia, come si dice, in "**CLASSI DI FREQUENZA**" (ad esempio in una rilevazione di altezze: $cm\ 150 \le h < cm\ 154,\ \ 154 \le h < 158,\ \ 158 \le h < 162,\ \ ecc.$), nel calcolo di una media si prende, per ciascuna classe, il cosiddetto "**VALORE CENTRALE**" **della classe**, ossia la **semisomma** (= la media) **delle estremità dell'intervallo**. Esempio:

$150 \le h < 154$	$154 \le h < 158$	$158 \le h < 162$	$162 \le h < 166$	$166 \le h < 170$
2	5	8	10	15
$170 \le h < 174$	$174 \le h < 178$	$178 \le h < 182$	$182 \le h < 186$	$186 \le h < 190$
9	7	5	2	1

$$M=\frac{152\cdot 2+156\cdot 5+160\cdot 8+164\cdot 10+168\cdot 15+172\cdot 9+176\cdot 7+180\cdot 5+184\cdot 2+188\cdot 1}{2+5+8+10+15+9+7+5+2+1}=\frac{10760}{64}=168,125$$

Nell'ultimo esempio anziché sommare un certo numero di addendi uguali abbiamo moltiplicato ciascun addendo per il numero di volte in cui questo andava considerato; abbiamo cioè fatto quella che, come vedremo poco più avanti, si chiama una "media PONDERATA".

E' evidente che questo metodo del "valore centrale" non fornisce come risultato la media "esatta", ma solo un valore approssimato della "vera" media.
La "vera" media, infatti, dovrebbe tenere conto di *tutti* i singoli valori osservati (che per comodità sono invece stati riuniti in classi);
ciascun singolo valore dovrebbe essere moltiplicato per la sua brava frequenza, questi prodotti sommati e infine questa somma divisa per il numero totale dei valori considerati.
L'approssimazione però in genere è molto buona …
Rinunciamo ad ulteriori approfondimenti, ma possiamo comunque fare un "esperimento" pratico.
Riprendiamo la tabella precedente ed entriamo nel dettaglio delle singole osservazioni:

$150 \le h < 154$		$154 \le h < 158$		$158 \le h < 162$		$162 \le h < 166$		$166 \le h < 170$	
150	0	154	0	158	1	162	2	166	4
151	0	155	1	159	2	163	2	167	4
152	1	156	2	160	3	164	3	168	3
153	1	157	2	161	2	165	3	169	4
	2		5		8		10		15

$170 \le h < 174$		$174 \le h < 178$		$178 \le h < 182$		$182 \le h < 186$		$186 \le h < 190$	
170	3	174	3	178	2	182	1	186	0
171	2	175	1	179	1	183	0	187	1
172	2	176	2	180	0	184	1	188	0
173	2	177	1	181	2	185	0	189	0
	9		7		5		2		1

Facendo, questa volta, la "vera" media si ottiene un valore vicino a $167,67$ quindi non molto differente da quello ricavato prima.

PROPRIETA' DEI VARI TIPI DI MEDIA

Se i nostri dati sono $x_1, x_2, \ldots, x_n$, e la loro media aritmetica è $M = \frac{x_1 + x_2 + \ldots + x_n}{n}$,
i loro "scarti" dalla media sono le differenze fra i dati stessi e la media:

$$\left.\begin{array}{c} x_1 - M \\ x_2 - M \\ \ldots \\ x_n - M \end{array}\right\} \textit{scarti dei dati dalla media}$$

Ecco la tabella delle altezze superate, in cm, da 7 atleti ad una gara dilettantistica di salto in alto:

180	180	184	184	184	190	200

Se ne calcoli la media, avrai

$$M = \frac{180+180+184+184+184+190+200}{7} = \frac{1302}{7} = 186$$

Ora scriviamo, sotto ciascuno dei dati, il suo scarto dalla media:

180	180	184	184	184	190	200
–6	–6	–2	–2	–2	+4	+14

Se a questo punto sommiamo algebricamente questi 7 scarti, avremo

$$-6-6-2-2-2+4+14=0$$

Il fatto che la somma algebrica degli scarti dalla media aritmetica sia 0 è del tutto generale.
In effetti, se $x_1, x_2, \ldots, x_n$ sono i dati,
e quindi $x_1 - M, x_2 - M, \ldots, x_n - M$ sono i loro scarti dalla media aritmetica, avremo

$$\textit{somma scarti} = (x_1 - M) + (x_2 - M) + \ldots + (x_n - M) = x_1 - M + x_2 - M + \ldots + x_n - M =$$

$$= x_1 + x_2 + \ldots + x_n - nM = x_1 + x_2 + \ldots + x_n - \cancel{n} \cdot \frac{x_1 + x_2 + \ldots + x_n}{\cancel{n}} = x_1 + x_2 + \ldots + x_n - (x_1 + x_2 + \ldots + x_n) = 0$$

PROPRIETÀ: **La somma degli scarti dei dati dalla media aritmetica dei dati stessi è sempre uguale a 0.**

Un'altra proprietà interessante della media aritmetica è la seguente:

PROPRIETÀ (che non dimostriamo; potresti però verificarla su di un esempio, tramite un foglio elettronico …)
La media aritmetica è quel valore rispetto al quale è minima la somma dei QUADRATI degli scarti.
Vale a dire, se io calcolo la somma dei quadrati degli scarti dalla media aritmetica,
questa somma sarà certamente minore di ciò che otterrei se, al posto degli scarti dalla media aritmetica M,
considerassi gli scarti da un qualsiasi altro valore a.
Schematicamente: se $x_1, x_2, \ldots, x_n$ sono i dati, e M è la loro media aritmetica, allora la quantità

$$(x_1 - a)^2 + (x_2 - a)^2 + \ldots + (x_n - a)^2$$

è minima nel caso $a = M$.

RIASSUNTO SCHEMATICO (INDICI DI POSIZIONE: le medie "ferme")

$\textbf{\textit{MEDIA ARITMETICA}} = M = \frac{x_1 + x_2 + \ldots + x_n}{n}$	"Conserva" la somma. EXCEL, OPENOFFICE CALC: = **MEDIA()**
$\textbf{\textit{MEDIA GEOMETRICA}} = M_G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$	"Conserva" il prodotto
$\textbf{\textit{MEDIA QUADRATICA}} = M_Q = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}}$	"Conserva" la somma dei quadrati
$\textbf{\textit{MEDIA ARMONICA}} = M_A = \frac{1}{\frac{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}{n}}$	"Conserva" la somma dei reciproci

B) MEDIE PONDERATE

1	9
2	14
3	25
4	22
5	10
	80

Avevamo già fatto qualche anticipazione. Riprendiamo il discorso.
Un test su 80 studenti universitari ha fatto registrare i punteggi della tabella qui a fianco (accanto a ciascun possibile punteggio, da 1 a 5, è stata annotata la relativa "frequenza", ossia il numero di studenti che hanno conseguito *quel* punteggio).
Qual è stata la media dei punteggi di questo gruppo di studenti?

$$M=\frac{\overbrace{1+...+1}^{9\ addendi}+\overbrace{2+...+2}^{14\ addendi}+\overbrace{3+...+3}^{25\ addendi}+\overbrace{4+...+4}^{22\ addendi}+\overbrace{5+...+5}^{10\ addendi}}{80}=\boxed{\frac{1\cdot 9+2\cdot 14+3\cdot 25+4\cdot 22+5\cdot 10}{80}}=$$

$$=\frac{9+28+75+88+50}{80}=\frac{250}{80}=3{,}125$$

Si dice "MEDIA PONDERATA" (o "MEDIA PESATA") una media nella quale ciascun dato viene moltiplicato per un fattore dato dalla sua frequenza assoluta, ossia è contato per un numero di volte uguale alla sua frequenza assoluta.

Nell'esempio di cui sopra, il dato "1" ha "peso" 9, il dato "2" ha "peso" 14, ecc.

Dunque, in generale, si ha,
per una media (aritmetica) ponderata,

$$\boxed{\begin{array}{l}\textit{Media}\\ \textit{aritmetica}=M=\dfrac{x_1\cdot f_1+x_2\cdot f_2+...+x_n\cdot f_n}{\underbrace{f_1+f_2+...+f_n}_{numero\ totale\ dei\ dati}}\\ \textit{ponderata}\end{array}}=\frac{\sum\limits_{k=1}^{n}x_k\cdot f_k}{\underbrace{\sum\limits_{k=1}^{n}f_k}_{\substack{numero\ totale\\ dei\ dati}}}$$

f_k = *frequenza assoluta del k-esimo dato* (*numero di volte in cui compare*)

Il simbolo $\sum\limits_{k=...}^{...}$ si chiama "simbolo di *sommatoria*".

Scrivere $\sum\limits_{k=1}^{n}x_k\cdot f_k$ significa che si vuole eseguire la somma di tanti addendi $x_k\cdot f_k$, dove k assume:

- il valore 1 (1° addendo),
- poi il valore 2 (2° addendo),
- eccetera,
- fino al valore n.

In questa interpretazione (ma vedi poi il riquadro sottostante), una "media ponderata" non differisce da una normalissima media. Semplicemente, visto che un dato si è presentato nella rilevazione più volte, lo si scrive, per comodità, una volta sola, *moltiplicandolo per la sua frequenza*, ossia per il numero di volte in cui compare.

❑ Verifica che se 26 persone hanno donato 5 euro e 14 persone 10 euro, la media dell'offerta è stata di euro 6,75

❑ Calcola la media del voto in condotta, nella pagella più recente, di tutti gli studenti della tua classe

Si parla di "media ponderata" anche quando si vogliono assegnare, ai dati $x_1, x_2, ..., x_n$, "PESI" DIVERSI in quanto i dati vengono ritenuti di diversa "importanza". La formula è la stessa, solo che al posto delle frequenze $f_1, f_2, ..., f_n$ ci sono i "pesi" $p_1, p_2, ..., p_n$ che si è deciso di attribuire ai vari dati; vedi qui a destra →

Un'ALTRA INTERPRETAZIONE ♥ della media ponderata

$$\boxed{M=\frac{x_1\cdot p_1+x_2\cdot p_2+...+x_n\cdot p_n}{\underbrace{p_1+p_2+...+p_n}_{somma\ dei\ pesi}}}$$

❑ Verifica che se i punteggi ottenuti da uno studente per i tre esercizi A, B, C sono stati rispettivamente 10, 8 e 7, e ai tre esercizi l'insegnante ha ritenuto di attribuire rispettivamente i "pesi" 2, 3 e 3, allora il punteggio dato dalla media ponderata risulta essere 8,125.
Verifica poi che se invece i "pesi" fossero 2, 3 e 4, quello studente otterrebbe come punteggio finale 8, mentre la media aritmetica "semplice" (= non ponderata) degli stessi punteggi è 8,333…

❑ Se in una doppia prova scritto+orale Anna è stata valutata rispettivamente 7 e 9, e la prof. intende assegnare peso 3 allo scritto e peso 1 all'orale per fare poi una media ponderata tramite la formula nell'ultimo riquadro, è come se Anna avesse preso i 4 voti 7, 7, 7 e 9 di cui fare poi la media "semplice" (= normale): verificalo.

C) LE MEDIE "LASCHE": MEDIANA E MODA (lasco = *allentato, molle, non teso*)

(Una media si dice "lasca" se potrebbe pure restare invariata, qualora cambiasse uno dei termini)

4	1
5	5
6	7
7	7
8	3
9	3
10	1
	27

In un test, i punteggi dei 27 studenti sono stati quelli riassunti dalla tabella qui a fianco (punteggio sulla colonna sinistra, frequenza assoluta di quel punteggio sulla colonna destra).

Se trascriviamo i punteggi uno a uno in ordine crescente, avremo

4	**5**	**5**	**5**	**5**	**5**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**8**	**8**	**8**	**9**	**9**	**9**	**10**

Consideriamo ora il punteggio che, nella striscia, occupa la **posizione centrale**:

4	**5**	**5**	**5**	**5**	**5**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**8**	**8**	**8**	**9**	**9**	**9**	**10**

Questo punteggio è 7. Diremo dunque che la "**mediana**" della distribuzione in esame è 7.
Se poi il "mostro" che ha preso 10 fosse stato assente, la striscia dei punteggi avrebbe contenuto 26 numeri:

4	**5**	**5**	**5**	**5**	**5**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**8**	**8**	**8**	**9**	**9**	**9**

In questo caso un numero che occupi la posizione centrale … non c'è. In tale eventualità (**numero pari di dati**)
si assume, convenzionalmente, come mediana la semisomma (= la media aritmetica)
dei due valori che stanno all'immediata sinistra e all'immediata destra della posizione centrale.

Nell'esempio considerato, quindi, la *mediana* sarebbe stata $\frac{6+7}{2}=6,5$.

Osserviamo che la *media aritmetica* dei punteggi della classe è
$\frac{181}{27}\approx 6,704$ con la presenza del "mostro", $\frac{171}{26}\approx 6,577$ supponendo il "mostro" assente. Dunque:

La **MEDIANA** è definita quando si ha un insieme di dati, disposti in ordine crescente.
Si tratta allora del **dato che "occupa il posto centrale"** ,
nel senso che **metà dei dati considerati sta a sinistra e metà a destra della mediana**.

Nel caso in cui il numero di questi dati sia pari, un "dato centrale" non esiste
e quindi, convenzionalmente, si assume come mediana la media aritmetica
fra i due dati che stanno immediatamente prima e immediatamente dopo, rispetto alla posizione centrale.

Qualora i dati non siano numerici, ma abbia comunque senso ordinarli
(livelli di istruzione, aggettivi che esprimono un gradimento …)
non ha senso pensare ad una "media" … ma a una *mediana*, in generale, sì.

Quando è possibile determinare sia la media aritmetica che la mediana, cioè con dati numerici,
la "mediana" ci dà un'informazione diversa rispetto alla "media".

Abbiamo già visto che la media aritmetica è quel valore che,
se venisse sostituito al posto di ciascuno di dati, ne lascerebbe inalterata la somma;
la mediana ci dice invece qual è il valore "centrale" della successione di dati, nel senso che,
se conosciamo la mediana, possiamo dire che un 50% dei dati è $\le$ e l'altro 50% è $\ge$ della mediana.

La mediana, rispetto alla media aritmetica,
è meno "sensibile" alla presenza di "dati anomali", cioè di dati "lontani dalla centralità".

Se nel precedente insieme di punteggi il punteggio più basso fosse stato "2" anziché "4",
la mediana non sarebbe variata, la media aritmetica sì.

PROPRIETÀ: **La mediana, se è un valore numerico,**
è quel valore rispetto al quale è minima la somma dei valori assoluti degli scarti.
Vale a dire, se io calcolo la somma dei valori assoluti degli scarti dalla mediana,
questa somma sarà certamente minore di ciò che otterrei se, al posto degli scarti dalla mediana,
considerassi gli scarti da un qualsiasi altro valore. *Verificalo empiricamente col foglio elettronico*!

E parliamo, infine, di **moda**.

Per MODA si intende **il dato che si è presentato con più frequenza**.

La moda potrebbe anche non essere unica!

Nell'esempio sopra considerato della classe col suo test,
ci sarebbero state due "mode": 6 e 7 (si parla in questo caso di distribuzione "bimodale").

punt	freq
4	1
5	5
6	**7**
7	**7**
8	3
9	3
10	1

Nel caso in cui le modalità sono suddivise in "classi", più che parlare di "moda"
è corretto parlare di "**CLASSE MODALE**" (= la classe con maggiore frequenza).

Quando abbia senso parlare tanto di media aritmetica, quanto di mediana, quanto di moda,
la moda ci dà un'informazione diversa rispetto alla media e alla mediana.

E osserviamo che nel caso in cui i dati siano di carattere qualitativo, e non abbia gran senso ordinarli,
non si può parlare né di media né di mediana, mentre la moda è comunque determinabile.

Se ad esempio un certo giorno di Agosto una gelateria ha venduto
12 granite al limone, 15 all'arancia e 7 al cedro, quel giorno la "moda" per le granite è stata "arancia",
senza che ovviamente si potesse parlare né di media aritmetica né di mediana.

UN'ESERCITAZIONE COL FOGLIO ELETTRONICO: MEDIE, CONTEGGI, ISTOGRAMMA

I pesi in kg dei 240 maschi maggiorenni di un villaggio sul fiume Yukon, in Alaska, sono stati registrati in un foglio elettronico.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	70,0	80,5	80,7	83,9	72,7	82,8	83,6	69,5	73,8	78,6	73,1	101,2	60,6	80,4	76,4	76,5
2	78,6	69,1	93,2	69,4	76,2	61,9	101,8	104,3	88,8	75,8	77,7	94,7	67,4	79,9	77,5	86,9
3	69,9	87,2	83,2	110,8	99,1	80,4	85,6	73,4	94,9	72,5	74,3	75,3	61,6	102,9	83,1	99,4
4	83,8	75,0	68,7	87,8	112,2	68,6	73,7	64,5	83,3	85,3	68,2	88,5	57,8	65,9	80,9	70,4
5	83,6	77,9	70,2	101,9	87,0	88,9	71,0	81,5	96,0	70,8	86,3	72,8	71,8	68,5	73,9	86,1
6	95,4	77,7	70,4	73,8	91,7	83,6	89,4	57,4	81,2	94,6	77,5	72,5	63,2	109,4	79,5	57,8
7	82,1	89,4	71,1	81,6	89,2	63,5	90,0	76,9	90,9	93,7	76,2	63,7	62,3	84,9	71,7	101,8
8	79,8	71,2	76,0	70,9	114,7	99,2	78,8	90,0	63,6	65,2	75,8	98,1	69,3	106,5	80,4	106,3
9	86,6	76,3	66,6	76,2	92,1	98,4	78,4	79,2	67,5	101,2	71,6	76,3	61,8	99,5	81,2	103,3
10	89,9	84,4	72,9	75,9	119,2	75,4	89,2	76,1	68,6	69,1	72,6	88,3	89,8	53,8	86,6	90,5
11	84,5	75,6	56,7	77,5	93,0	101,9	80,3	67,0	72,2	109,7	80,2	78,4	82,3	66,1	85,1	70,5
12	98,0	85,2	64,9	80,9	98,4	103,5	75,1	82,7	59,6	66,2	79,8	99,3	91,3	72,2	93,7	97,8
13	81,9	76,3	67,2	68,0	96,2	78,4	90,8	79,6	67,1	71,1	80,3	67,7	91,9	77,3	84,0	60,9
14	93,1	96,5	73,7	90,9	56,8	69,1	92,2	73,4	60,4	90,8	81,5	70,1	81,5	76,0	72,7	91,2
15	85,2	80,2	80,6	83,5	74,5	57,0	88,0	71,6	72,9	77,8	75,0	90,4	98,0	105,6	68,7	84,7

a) Determinare il peso minimo e il peso massimo
b) determinare media e mediana dei pesi
c) contare il numero di persone il cui peso rientra nella fascia da 50 kg a 60 kg (esclusi), da 60 a 70, ecc., e tracciare un istogramma;
d) determinare la media dei pesi suddivisi in "classi", assegnando a ogni classe il peso centrale fra i suoi due estremi, e ricalcolare la media per confrontarla con la media reale.

a) Possiamo posizionarci in una cella libera qualsiasi, ad esempio la A18, e digitare
= min(A1 : P15)

BELLO
Osserviamo che dopo aver digitato **= min(**
se **clicchiamo sulla cella A1** il foglio inserirà **automaticamente** nella formula il riferimento ad A1.
A questo punto digiteremo i "due punti": dopodiché potremo cliccare su P15 e infine chiudere la parentesi.
Comoda alternativa: si può digitare **= min(** poi **TRASCINARE** il mouse sul rettangolo da A1 a P15.

L'effetto finale, in A18, sarà

17	
18	53,8
19	

Allo stesso modo, in B18 inseriremo la formula
= max(A1 : P15)
ottenendo

17		
18	53,8	119,2
19		

Naturalmente, sarà opportuno inserire in celle adiacenti, **stringhe adeguate** che ci aiutino a ricordare il significato dei numeri ottenuti: ad es.

18	53,8	119,2
19	min	max

b) Digitiamo, ad esempio in C18 e in D18, le formule
= media(A1 : P15)
e rispettivamente
= mediana(A1 : P15) …

… nonché, in C19 e D19, le **stringhe opportune**, con l'effetto seguente:

18	53,8	119,2	80,8	79,7
19	min	max	media	mediana

c) Digitiamo in E18:
= conta.se(A1 : P15;" >= 50") – conta.se(A1 : P15;" >= 60")
e ci comparirà così, in E18, il numero di dati compresi fra 50 (incluso) e 60 (escluso):

	A	B	C	D	E
17					50-60
18	53,8	119,2	80,8	79,7	8
19	min	max	media	mediana	

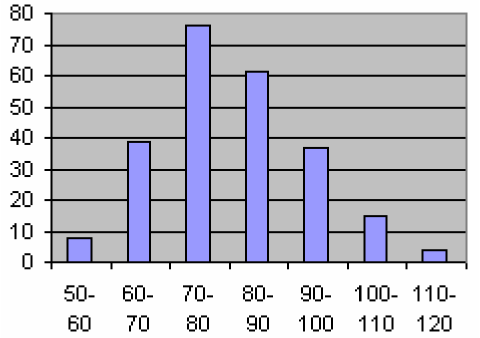
Procediamo in modo analogo sulle celle F18, G 18 … fino a K18:

	A	B	C	D	E	F	G	H	I	J	K
17					50-60	60-70	70-80	80-90	90-100	100-110	110-120
18	53,8	119,2	80,8	79,7	8	38	77	61	37	15	4
19	min	max	media	mediana							

Ora possiamo **selezionare, trascinando col mouse, il rettangolo di celle E17:K18**

17					50-60	60-70	70-80	80-90	90-100	100-110	110-120
18	53,8	119,2	80,8	79,7	8	38	77	61	37	15	4
19	min	max	media	mediana							

… **cliccare su** e, con qualche passaggio molto intuitivo, **ottenere finalmente l'istogramma**:



d) **Digitiamo, accanto alle frequenze delle classi, il "valore centrale" della classe …**

17					50-60	60-70	70-80	80-90	90-100	100-110	110-120
18	53,8	119,2	80,8	79,7	8	38	77	61	37	15	4
19	min	max	media	mediana	55	65	75	85	95	105	115

… e avviamoci ora a calcolare una MEDIA PONDERATA. In E20 inseriremo la formula
= E18 * E19

Arial 10 G C S % 000 €

E20 =E18*E19

	A	B	C	D	E	F	G	H	I	J	K
16											
17					50-60	60-70	70-80	80-90	90-100	100-110	110-120
18	53,8	119,2	80,8	79,7	8	38	77	61	37	15	4
19	min	max	media	mediana	55	65	75	85	95	105	115
20					440						

che poi incolleremo, trascinando il quadratino in basso a destra della cella, sulle celle limitrofe F20 … K20

17					50-60	60-70	70-80	80-90	90-100	100-110	110-120
18	53,8	119,2	80,8	79,7	8	38	77	61	37	15	4
19	min	max	media	mediana	55	65	75	85	95	105	115
20					440	2470	5775	5185	3515	1575	460

Ora in L18 e in L20 calcoliamo la **somma delle frequenze assolute** e, risp., la **somma dei prodotti** …

17					50-60	60-70	70-80	80-90	90-100	100-110	110-120	
18	53,8	119,2	80,8	79,7	8	38	77	61	37	15	4	240
19	min	max	media	mediana	55	65	75	85	95	105	115	
20					440	2470	5775	5185	3515	1575	460	19420

… per terminare in bellezza con la formula, inserita in L21:
= L20 / L18
che ci dà la "**MEDIA PER CLASSI**", molto vicina, come possiamo osservare, alla vera media precedentemente determinata.

15	4	240
105	115	
1575	460	19420
		80,917

9. GLI INDICI DI DISPERSIONE

Ritorniamo alla congettura del professor Curiosi (pagina 354) riguardo alle sue nuove classi I A e I B.

Il professore aveva avuto l'impressione, da una iniziale sommaria conoscenza,
che in una di esse gli studenti fossero "meno omogenei nella preparazione":
che ci fosse, insomma, un gruppo abbastanza nutrito di allievi molto bravi e un altro gruppo sostanzioso di scarsi.
Nell'altra classe la situazione gli era sembrata diversa, più equilibrata.

Dopodiché il professore aveva somministrato alle due classi il medesimo test di ingresso,
che aveva fatto registrare i punteggi seguenti (M = media):

I A 51 62 42 58 60 68 61 68 64 70 71 60 51 62 41 51 36 47 58 73 37 54 63 65 ($M \approx 57{,}2$)

I B 45 48 51 63 51 60 29 52 47 41 52 50 56 62 57 70 55 64 59 55 67 ($M = 54$)

Ci domandiamo ora:
esisterà un indicatore statistico adeguato a valutare se il test effettuato conferma l'impressione iniziale?

Un primo indicatore di "dispersione" (= di "sparpagliamento" dei dati) potrebbe essere
la *differenza fra il dato massimo e il dato minimo* in ciascuna delle due classi.
Vediamo che

- per la I A questa differenza, detta in statistica "**campo di variabilità**", vale $73-36=37$
- mentre in I B vale $70-29=41$.

$$\boxed{\textit{campo di variabilità} = \textit{dato massimo} - \textit{dato minimo} = x_{\mathrm{MAX}} - x_{\mathrm{min}}}$$

A giudicare dal "campo di variabilità", sembrerebbero quindi più disomogenee le prestazioni della I B …

… tuttavia, va osservato che il "campo di variabilità" tiene conto di DUE SOLI valori (quelli estremi)
mentre non risente per nulla di tutti i valori intermedi … la presenza, nella classe, anche di *un singolo*
caso isolato di alunno molto bravo o molto poco preparato potrebbe allora condizionarlo pesantemente.
Le prestazioni della "massa" degli allievi non influiscono in alcun modo sul calcolo di questo indicatore!

Riflettiamo. Quello che veramente ci interessa è di investigare
in quale delle due classi i valori "sono mediamente più lontani dalla media aritmetica".

Potremmo allora pensare, per ciascuna classe, di elencare tutti gli "scarti dalla media".

I A		51	62	42	58	…
$M \approx 57{,}2$	Scarti	–6,2	+4,8	–15,2	+0,8	…
I B		45	48	51	63	…
$M = 54$	Scarti	–9	–6	–3	+9	…

Questo sarebbe un buon inizio, ma poi?
Se ora andassimo a calcolare la *media aritmetica di questi scarti*, per entrambe le classi *otterremmo 0*!
E certo! Come sappiamo, infatti, la somma algebrica degli scarti dalla media aritmetica è sempre 0.

Sorge allora l'idea di calcolare la media aritmetica … *non* degli scarti, ma del VALORE ASSOLUTO di questi.
Tale media si dice "**scarto medio**" o (più correttamente) "**scarto assoluto medio**".

$$\boxed{\begin{array}{l}\textit{scarto medio} = \textit{media aritmetica dei valori assoluti degli scarti dalla media aritmetica} = \\ \quad = \dfrac{|x_1 - M| + |x_2 - M| + \ldots + |x_n - M|}{n} \; (\textit{è più corretto dire}: \textit{"scarto assoluto medio"})\end{array}}$$

Così facendo, otteniamo (verificalo con un foglio elettronico!) $\textit{scarto}(\mathrm{IA}) \approx 8{,}74$; $\textit{scarto}(\mathrm{IB}) \approx 7{,}05$.

Vediamo di trarre qualche conclusione.
Per la I A, abbiamo ottenuto $\textit{campo di variabilità}\,(\mathrm{IA}) = 37$; $\textit{scarto assoluto medio}\,(\mathrm{IA}) \approx 8{,}74$
e per la I B $\textit{campo di variabilità}\,(\mathrm{IB}) = 41$; $\textit{scarto assoluto medio}\,(\mathrm{IB}) \approx 7{,}05$

♪ **La I A ha uno scarto assoluto medio maggiore**:
i punteggi sono mediamente più lontani, in questa classe, dalla media aritmetica della classe,
segno della presenza "importante" di fasce di allievi che si allontanano alquanto dalla media

♫ D'altra parte, il **campo di variabilità è maggiore per la I B**:
di ciò è responsabile il povero alunno che, purtroppo, ha conseguito un punteggio bassissimo (29 punti).

Anziché fare la media dei valori assoluti degli scarti,
avremmo potuto anche *elevare ciascuno scarto al quadrato*, ottenendo così un valore certamente positivo,
per poi fare la media aritmetica dei QUADRATI degli scarti (detta "varianza").

$$varianza = media\ aritmetica\ dei\ quadrati\ degli\ scarti\ dalla\ media\ aritmetica = \\ = \frac{(x_1 - M)^2 + (x_2 - M)^2 + \ldots + (x_n - M)^2}{n}$$

In questo modo avremmo avuto $varianza\,(\text{I A}) \approx 109{,}8$; $varianza\,(\text{I B}) \approx 83{,}2$

Varianza *maggiore* comporta *maggiore dispersione dei dati rispetto alla media della popolazione*:
la varianza, in accordo con lo scarto assoluto medio, indica dunque nella I A la classe più disomogenea.

Son pronto a scommettere che la "varianza" ti appare d'istinto più "antipatica" rispetto allo "scarto assoluto medio", che a prima vista sembra assai più semplice e più "spontaneo" da usare, come indice di dispersione.
Tuttavia, ti segnalo che nella pratica si preferisce invece utilizzare la "varianza", e ancora di più la sua radice quadrata che è chiamata "scarto quadratico medio", anziché lo "scarto assoluto medio".

I motivi per cui la "varianza" ha un rilievo speciale in statistica sono parecchi.
Qui ci limitiamo a citarne soltanto due.

1) La **varianza** $\frac{(x_1 - M)^2 + (x_2 - M)^2 + \ldots + (x_n - M)^2}{n}$
   è legata alla media aritmetica in modo assai peculiare.
   Infatti si può dimostrare che essa è sempre inferiore a qualsivoglia analoga quantità
   $\frac{(x_1 - a)^2 + (x_2 - a)^2 + \ldots + (x_n - a)^2}{n}$ nella quale gli scarti vengano calcolati,
   invece che rispetto alla media aritmetica *M*, rispetto ad un altro qualsiasi valore *a*.

   Lo "**scarto assoluto medio**" dal canto suo **si ricollega piuttosto ad *un altro* indice** di posizione centrale:
   la mediana. In effetti la quantità $\frac{|x_1 - a| + |x_2 - a| + \ldots + |x_n - a|}{n}$ è minima (come si potrebbe dimostrare)
   quando il valore *a* è la *mediana*, NON la media aritmetica dei dati $x_1, x_2, \ldots, x_n$.

2) **La varianza è il quadrato dello "scarto quadratico medio"**, di cui andiamo a parlare qui di seguito,
   e lo "scarto quadratico medio" ha un'importanza colossale in svariate questioni,
   come la teoria degli errori di misura.

Lo "**scarto quadratico medio**" o "**deviazione standard**" è la radice quadrata della varianza:

$$scarto\ quadratico\ medio\ o\ deviazione\ standard = radice\ quadrata\ della\ varianza = \\ = \sqrt{\frac{(x_1 - M)^2 + (x_2 - M)^2 + \ldots + (x_n - M)^2}{n}} = media\ quadratica\ degli\ scarti$$

Lo scarto quadratico medio viene generalmente indicato con σ ("sigma"), e la varianza con σ^2.
Nell'esempio precedentemente considerato dei punteggi delle due classi I A e I B, si ha:

$$\sigma^2(\text{I A}) = \frac{(x_1 - M)^2 + (x_2 - M)^2 + \ldots + (x_n - M)^2}{n} \approx \frac{(51 - 57{,}2)^2 + (62 - 57{,}2)^2 + \ldots + (65 - 57{,}2)^2}{24} \approx 109{,}8$$

$$\sigma^2(\text{I B}) = \frac{(x_1 - M)^2 + (x_2 - M)^2 + \ldots + (x_n - M)^2}{n} = \frac{(45 - 54)^2 + (48 - 54)^2 + \ldots + (67 - 54)^2}{21} \approx 83{,}2$$

da cui $\sigma(\text{I A}) = \sqrt{\sigma^2(\text{I A})} \approx 10{,}5$; $\sigma(\text{I B}) = \sqrt{\sigma^2(\text{I B})} \approx 9{,}1$

Se i dati provengono da una tabella con le frequenze, evidentemente sarà, dette f_i le frequenze (assolute):

$$\sigma = \sqrt{\frac{(x_1 - M)^2 f_1 + (x_2 - M)^2 f_2 + \ldots + (x_p - M)^2 f_p}{f_1 + f_2 + \ldots + f_p}}$$

Le ragioni per cui lo scarto quadratico medio è preferito alla varianza sono sostanzialmente due.

1) La prima è che, se i dati sono, ad esempio, dei metri, la "varianza" sarebbe espressa in "metri quadrati", e lo scarto quadratico medio invece ancora in metri. Insomma,
   lo scarto quadratico medio ha il pregio di avere la stessa unità di misura dei dati dei quali proviene.

2) La seconda ragione è il **ruolo cruciale dello scarto quadratico medio nella cosiddetta "gaussiana"**, alla quale accenneremo parlando, più avanti, di "**errori di misura**".

Per il calcolo dello scarto quadratico medio, anziché la formula $\sigma=\sqrt{\frac{(x_1-M)^2+(x_2-M)^2+...+(x_n-M)^2}{n}}$,

si può anche utilizzare una formula equivalente, più comoda, che è $\boxed{\sigma=\sqrt{\frac{x_1^2+x_2^2+...+x_n^2}{n}-M^2}}$.

Per *confrontare* due distribuzioni in quanto alla loro "variabilità", alla loro "dispersione", si utilizza un indice che è detto "coefficiente di variazione" (di solito espresso come percentuale, non calcolabile se la media dei dati è 0, e comunque poco significativo quando la media dei dati è vicina a 0):

$$\boxed{coefficiente\ di\ variazione=\frac{scarto\ quadratico\ medio}{|media\ \ aritmetica|}=\frac{\sigma}{|M|}}$$

NOTA - **Il coefficiente di variazione**, essendo il rapporto fra due quantità, σ e $|M|$, che sono espresse nella stessa unità di misura, **è un numero puro, senza unità di misura** (si dice che è "adimensionale").

Ad esempio, se si vanno a misurare i pesi dei bambini nati in un certo periodo in un grande ospedale, e simultaneamente i pesi delle loro mamme, si osserverà certamente una deviazione standard molto inferiore nell'insieme dei bambini … Per forza! Infatti i bambini appena nati pesano soltanto due-tre o quattro chili … quindi anche gli scarti dalla media dei loro pesi saranno piccolini!!! Volendo confrontare le due "variabilità" (quella dei pesi dei neonati con quella dei pesi delle mamme) si farà ricorso allora al coeff. di variazione.

RIASSUNTO SCHEMATICO (INDICI DI DISPERSIONE)

Indicatori di "DISPERSIONE" o di "VARIABILITÀ": ci dicono QUANTO, GLOBALMENTE, I DATI SONO LONTANI DALLA LORO MEDIA ARITMETICA *M*.

Ogni indicatore di dispersione ha la proprietà di essere maggiore quando i dati si allontanano maggiormente, nel loro complesso, dalla centralità.

CAMPO DI VARIABILITA' $=$ $=dato\ massimo-dato\ minimo=$ $=x_{MAX}-x_{min}$	E' un indicatore piuttosto "grezzo", perché dipende esclusivamente dai due valori estremi ignorando quelli intermedi	EXCEL, OPENOFFICE: **MAX()–MIN()**
SCARTO MEDIO = SCARTO ASSOLUTO MEDIO $=$ $=m.\ aritm.\ dei\ val.\ ass.\ degli\ scarti\ dalla\ m.\ aritm.=$ $=\frac{\|x_1-M\|+\|x_2-M\|+...+\|x_n-M\|}{n}$	Sarebbe minimo qualora al posto della media *M* ci fosse, nella formula, la mediana	EXCEL, OPENOFFICE: **MEDIA.DEV()**
VARIANZA $=$ $=\sigma^2=m.\ aritm.\ dei\ quadr.\ degli\ scarti\ dalla\ m.\ aritm.=$ $=\frac{(x_1-M)^2+(x_2-M)^2+...+(x_n-M)^2}{n}$	Ha il difetto di non essere espressa nella stessa unità di misura dei dati	EXCEL, OPENOFFICE: **VAR.POP()** (NOTA)
SCARTO QUADR. MEDIO o DEVIAZ. STANDARD $=$ $=\sigma=radice\ quadrata\ della\ varianza=$ $=\sqrt{\frac{(x_1-M)^2+(x_2-M)^2+...+(x_n-M)^2}{n}}=$ $=media\ quadratica\ degli\ scarti$ Comodissima formula alternativa: $\sigma=\sqrt{\frac{x_1^2+x_2^2+...+x_n^2}{n}-M^2}$	E' l'indicatore di dispersione più utilizzato in statistica; è espresso nella stessa unità di misura dei dati, e ha un'importanza decisiva nella teoria degli errori di misura, e, in generale, nelle distribuzioni che tendono a identificarsi con la cosiddetta "gaussiana"	EXCEL, OPENOFFICE: **DEV.ST.POP()** (NOTA)
COEFF. DI VARIAZ. $=\frac{scarto\ quadratico\ medio}{\|media\ \ aritmetica\|}=\frac{\sigma}{\|M\|}$	E' un numero puro, senza unità di misura, ottimo per confrontare fra loro distribuzioni differenti.	

NOTA su alcune funzioni statistiche nel foglio elettronico

$$\mathrm{VAR.POP}=\frac{(x_1-M)^2+(x_2-M)^2+...+(x_n-M)^2}{n},\quad \mathrm{VAR}=\frac{(x_1-M)^2+(x_2-M)^2+...+(x_n-M)^2}{n-1}$$

VAR è dunque, per il foglio elettronico, la cosiddetta "**varianza corretta**", ossia un indicatore statistico che, calcolato su di un campione, permette di stimare meglio la varianza incognita dell'intera popolazione. La "varianza corretta" e l'analoga "deviazione standard corretta" si utilizzano quindi in statistica inferenziale … questo tuttavia è un discorso che, se affrontato seriamente, presenta grande interesse ma anche una certa difficoltà.

TM	TF	BM	BF
96,3	96,4	58	57
96,7	96,7	63	57
96,9	96,8	64	59
97	97,2	64	61
97,1	97,2	64	61
97,1	97,4	65	62
97,1	97,6	66	62
97,2	97,7	66	64
97,3	97,7	67	64
97,4	97,8	67	64
97,4	97,8	68	65
97,4	97,8	68	65
97,4	97,9	68	66
97,5	97,9	69	66
97,5	97,9	69	68
97,6	98	70	68
97,6	98	70	68
97,6	98	70	69
97,7	98	70	69
97,8	98	70	69
97,8	98,1	70	69
97,8	98,2	71	70
97,8	98,2	71	71
97,9	98,2	71	72
97,9	98,2	71	73
98	98,2	71	73
98	98,2	72	73
98	98,3	72	73
98	98,3	72	73
98	98,3	72	74
98	98,4	73	74
98,1	98,4	73	75
98,1	98,4	73	76
98,2	98,4	73	76
98,2	98,4	73	77
98,2	98,5	74	77
98,2	98,6	74	77
98,3	98,6	74	77
98,3	98,6	74	77
98,4	98,6	75	78
98,4	98,7	75	78
98,4	98,7	75	78
98,4	98,7	75	79
98,5	98,7	76	79
98,5	98,7	77	79
98,6	98,7	77	79
98,6	98,8	78	79
98,6	98,8	78	79
98,6	98,8	78	80
98,6	98,8	78	80
98,6	98,8	78	81
98,7	98,8	78	81
98,7	98,8	78	81
98,8	98,9	79	82
98,8	99	80	82
98,8	99	80	83
98,9	99,1	81	83
99	99,1	81	84
99	99,2	82	84
99	99,2	82	84
99,1	99,3	82	85
99,2	99,4	83	86
99,3	99,9	83	87
99,4	100	84	89
99,5	100,8	86	89

I dati qui a sinistra sono tratti dal
Journal of the American Medical Association, vol. 268.

Di 130 soggetti, 65 uomini e 65 donne,
rappresentanti un campione casuale della popolazione locale,
sono stati misurati

- la temperatura corporea, in gradi Fahrenheit,
- e il numero di battiti cardiaci al minuto.

Utilizza un foglio elettronico per calcolare, di ciascuna colonna,

- la media
- lo scarto quadratico medio o deviazione standard
- lo scarto quadratico medio "corretto"
- il coefficiente di variazione
  (prendi lo sc. q. m. "non corretto" per determinarlo)

Le risposte sono qui in fondo alla pagina, capovolte,
ma tu guardale solo alla fine!

Per trovare altri gruppi di dati reali "grezzi" su cui lavorare,
puoi ad esempio consultare le pagine web
www.amstat.org/publications/jse/jse_data_archive.htm
e
http://www2.stetson.edu/~jrasp/data.htm

LA STATISTICA di *Trilussa*

Sai ched'è la statistica? È 'na cosa
che serve pe' fa' un conto in generale
de la gente che nasce, che sta male,
che more, che va in carcere e che spósa.
Ma pe' me la statistica curiosa
è dove c'entra la percentuale,
pe' via che, lì, la media è sempre eguale
puro co' la persona bisognosa.
Me spiego: da li conti che se fanno
seconno le statistiche d'adesso
risurta che te tocca un pollo all'anno:
e, se nun entra ne le spese tue,
t'entra ne la statistica lo stesso
perchè c'è un antro che ne magna due.

In tutti i casi seguenti
c'è chi mangia 0 polli e chi ne mangia di più:
secondo te, quali situazioni sono
più equilibrate, meno ingiuste?

Prova a calcolare media,
scarto quadratico medio,
coefficiente di variazione …

2 persone: 0 polli, 2 polli
3 persone: 0 1 2
5 persone: 0 1 1 1 2
6 persone: 0 0 1 1 2 2
4 persone: 0 0 0 4
6 persone: 0 0 0 1 1 4
3 persone: 0 2 4
5 persone: 0 3 3 3 6

media	74,15384615	73,36923077	98,39384615	98,10461538
devstandard	8,042637855	5,829815225	0,737746449	0,693359884
coeffvariaz	0,108458809	0,079458585	0,007497892	0,007067556
devstandardcorr	8,105227421	5,875184122	0,743487753	0,698755762

10. - ESERCIZI

Sugli INDICI DI POSIZIONE
(risposte a pag. 423)

A) MEDIA ARITMETICA, MEDIANA, MODA

1) I voti di una verifica sono stati i seguenti. Quanto valgono la media, la moda e la mediana?

7	7,5	8,5	5,5	6	8	7,5	6,5	5,5	8	4
6	4	6,5	6,5	7,5	9	5	4,5	7	8,5	9

2) "Ti ricordi quante erano mediamente le tue ore di studio pomeridiane, l'anno scorso"?
Fu fatta questa domanda a un gruppo di alunni di Prima Liceo, ed essi risposero così:

1	0,5	2	3	1,5	3	3	2	1	1	1,5	3	2,5	2	1,5	1,5
2	1,5	4	1	1,5	2,5	2	0,5	3	1,5	3	2	1	1,5	2	

Determina media aritmetica, mediana e moda di questi dati.

3) L'altezza media di 5 pallavoliste professioniste è di m 1,78. Quanto dovrebbe essere alta, al minimo, una sesta atleta, per far sì che la media raggiunga almeno metri 1,80?

4) Immaginiamo di suddividere un insieme di dati in due parti.
La media (aritmetica) generale coinciderà con la media delle due medie? Tu cosa ne dici?

5) Se in una regione un certo partito ha avuto il 26% dei consensi fra gli aventi diritto al voto e nella regione limitrofa solo il 16%, riunendo insieme le due regioni che percentuale si otterrebbe?

6) Aldo e Bruno si sono allenati sullo stesso percorso podistico.
Aldo, tutti i giorni feriali della settimana;
Bruno un giorno in meno perché ha perso, per un impegno, un allenamento.
I tempi di percorrenza sono stati quelli in tabella.
Aldo ha dunque fatto il "record" con 24' 19".
E riguardo alle medie delle prestazioni, chi è stato il più veloce?

	Aldo	*Bruno*
Lunedì	25' 54"	26' 04"
Martedì	24' 45"	25' 55"
Mercoledì	25' 58"	24' 35"
Giovedì	26' 24"	
Venerdì	24' 19"	25' 18"

B) MEDIA ARITMETICA PER CLASSI

7) Numero di giorni in cui un libro è stato trattenuto in prestito dagli utenti di una biblioteca scolastica.

7	15	4	10	21	11	9	5	23	28	18	12	15	14	13	22	19	20	7	15	18	12	18
13	12	9	7	9	15	14	10	5	14	16	18	19	21	22	3	16	19	7	26	15	17	12

a) Calcola la media
b) Ricalcola la media dopo aver raggruppato i dati in intervalli ("classi di frequenza") di 7 giorni (da 1 a 7; da 8 a 14; da 15 a 21; da 22 a 28 giorni).
Ricorda che, quando i dati sono suddivisi in classi, il valore che si attribuisce a ciascuna classe è la *semisomma* degli estremi dell'intervallo. Nel nostro esempio, la classe "da 1 a 7 " ha frequenza 8 (se fai il conteggio, vedrai che sono 8 gli utenti che hanno trattenuto il libro da 1 a 7 giorni); bene, allora nel calcolo della media per classi si moltiplicherà per 8 il valore centrale della classe ossia (1+7)/2=4.
La media *per classi* trovata differisce di molto dalla media "normale"?

8) Ecco qui di seguito la spesa registrata da una delle casse di un supermercato per 100 persone consecutive.

13,65	83,20	12,45	7,05	123,40	22,10	75,25	32,15	95,50	135,45
67,25	31,20	14,70	68,80	72,15	25,50	151,15	39,80	48,45	35,50
19,95	64,10	84,15	30,95	22,85	36,70	61,85	14,40	23,55	82,15
23,60	35,90	41,10	52,80	7,65	23,30	129,25	57,15	18,10	97,25
32,10	45,45	59,65	34,45	47,10	22,50	33,90	97,05	24,55	149,60
19,95	42,50	76,20	84,15	24,5	32,45	23,20	60,35	43,80	25,55
28,10	39,10	45,50	71,05	30,90	28,15	16,25	77,10	36,35	42,15
254,50	100,90	93,15	27,75	9,90	38,45	26,50	20,95	30,40	55,00
36,10	28,90	73,85	48,70	98,95	29,40	35,15	44,20	36,30	22,55
19,95	20,75	66,80	51,05	34,70	40,80	46,05	63,25	89,10	33,80

Con un foglio elettronico:
a) individua la spesa minima e la massima b) calcola la spesa media individuale
c) ordina i dati d) suddividi i dati in classi la cui ampiezza sia 10 euro
e) ricalcola la media "per classi" (si conta il numero di clienti la cui spesa rientra in una data classe, e si attribuisce come spesa a ciascuno di quei clienti il valore centrale della classe).
La media "esatta" e la media "per classi" così determinate differiscono di molto?

9) I prezzi di un articolo, rilevati in un gruppo di 72 esercizi commerciali italiani, sono stati quelli riportati nella tabella qui a fianco.

Se ti viene richiesta la media dei prezzi in questo insieme di negozi, sei in grado di calcolarla?

Prezzo (in euro)	*Numero punti vendita*
$8,00 \le p < 8,50$	2
$8,50 \le p < 9,00$	5
$9,00 \le p < 9,50$	15
$9,50 \le p < 10,00$	24
$10,00 \le p < 10,50$	18
$10,50 \le p < 11,00$	7
$11,00 \le p < 11,50$	1

C) MEDIA ARITMETICA PONDERATA (NELLE DUE INTERPRETAZIONI)

10) I genitori modificano la paghetta settimanale del figlio a seconda del comportamento, e dei voti a scuola. L'anno passato il ragazzo ha avuto: per 25 settimane 20 euro a settimana, per 18 settimane 10 euro e per le rimanenti 9 settimane … 0 euro.
Quale è stata la paghetta settimanale media?

11) In occasione del pensionamento di un collega, viene fatta una colletta per acquistare un regalo e 5 partecipanti donano 20 euro ciascuno, 18 contribuiscono con 10 euro, i rimanenti 23 ci mettono 5 euro.
Calcola media, mediana e moda delle offerte.

12) In una colonia estiva i ragazzi hanno le età in tabella.

Età	Numero ragazzi
10	25
11	28
12	31
13	18

Qual è l'età media? L'età mediana? La moda delle età?

13) Un grosso complesso residenziale ha appartamenti di varia conformazione. La tabella indica quanti fra gli appartamenti hanno *n* vani.
Determina la media, la mediana, la moda del numero di vani.

	n
4	2
15	3
25	4
20	5
5	6
2	7

14) a) Una giovane insegnante con poca esperienza decide di assegnare un punteggio da 0 a 10 a ciascuna delle 5 parti A, B, C, D, E in cui si articola una prova scritta. Corregge i primi 3 elaborati e annota i vari giudizi parziali in una griglia, ripromettendosi poi di fare la media su ogni riga:

Studente	A	B	C	D	E
Paolo	8	7	7	7	8
Serena	7	8	8	8	9
Martina	8	9	9	6	7

Le tre correzioni effettuate, però, inducono la professoressa ad un ripensamento, perché fanno emergere con chiarezza che non sarebbe corretto considerare i 5 quesiti equivalenti fra loro: alcuni infatti risultano essere ben più impegnativi di altri.
Decide allora di "pesare" in modo diverso le differenti sezioni, e attribuisce i pesi in questo modo:

Parte	A	B	C	D	E
Peso	1	1,5	0,8	2	0,5

Come verranno valutati dunque Paolo, Serena e Martina, se quello che si vuole è un voto finale da 0 a 10?

E se si desidera un voto finale dal 2 al 10?

b) Realizza un **foglio elettronico** in cui un insegnante possa inserire, per una verifica con 5 esercizi: il punteggio (da 0 a *puntmax*) acquisito in ciascun esercizio; il "peso" attribuito a ogni esercizio; il voto minimo e il voto massimo previsti. Chiaramente, ne dovrà uscire il voto assegnato.

15) Una ditta che vuole assumere tecnici specializzati valuta per ciascun candidato: il curriculum iniziale; gli esiti di un esame scritto; gli esiti di un colloquio orale. I punteggi sono in decimi; tuttavia, si è deciso di assegnare peso 1,5 all'esame scritto, ritenuto più indicativo degli altri due elementi di giudizio, mentre sia al curriculum che all'esame orale verrà attribuito peso 1.

Ciò premesso, con un foglio elettronico determina le medie ponderate degli 8 candidati. Qual è la minima fra queste?

Candidato	Curriculum	Scritto	Orale
A	5	7	7
B	4	6	8
C	7	5	5
D	2	8	9
E	8	7	8
F	4	6	6
G	3	5	4
H	7	9	8

D) ALTRI ESERCIZI SU MEDIA ARITMETICA, MEDIANA E MODA

16) Nella classe Seconda A, che ha 22 allievi, la media dei punteggi di un test è stata 7,25. Nella Seconda B gli allievi sono 28 e la media dei punteggi dello stesso test è risultata essere 7,8. E' possibile, con questi dati, calcolare esattamente la media complessiva, ossia la media dei punteggi ottenibili mettendo insieme in un unico gruppo tutti gli studenti di entrambe le classi?

17) Una squadra di basket, con 10 giocatori fra titolari e riserve, ha la sua brava distribuzione di altezze. Se il giocatore più alto (m 1,98) viene venduto ad un'altra squadra e il suo posto viene preso da un giocatore alto addirittura m 2,04, cambieranno media aritmetica e mediana delle altezze? Supponendo di suddividere le altezze in intervalli di 5 cm, cambierà la classe modale?

18) Al termine della frequentazione di una scuola privata, viene rilasciato un diploma comprensivo di valutazione finale che può essere un numero intero da 6 a 10. Se nella storia di quell'istituto scolastico fino ad oggi il 20% dei diplomati è uscito col 6, il 40% col 7, il 22% con l'8, il 12% col 9 e il 6% col 10, quale è stata la media di tutti i voti?

19) Con gli esiti del "questionario del curiosone" (pag. 360), calcola la *media* dei dati, laddove abbia significato, ossia per 1), 2), 3), 4), 5), 6), 7), 11), 12), 13), 15).
Per 1), ripeti poi il calcolo della media suddividendo le altezze in intervalli ("classi") di 3 cm e assumendo come valore, per ogni classe, il "punto di mezzo" di quella classe.
Ad esempio, se una delle classi è formata dalle altezze di cm 170-171-172, e gli alunni di questa fascia sono 5 con altezze date da 170, 170, 171, 171, 172, allora il valore centrale è 171 e nella media generale, anziché la somma 170+170+171+171+172, a numeratore comparirà 171 moltiplicato per 5.
Confronta il valore così ottenuto con la media calcolata precedentemente.

20) Con gli esiti del "questionario del curiosone", calcola la *mediana* dei dati, laddove abbia significato, ossia per 1), 2), 3), 4), 5), 6), 7), 10), 11), 12), 13), 15)

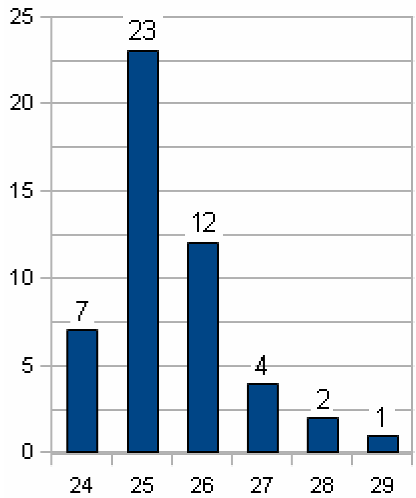
21) Con gli esiti del "questionario del curiosone", calcola la *moda* dei dati, laddove abbia significato, ossia per 1), 2), 3), 4), 5), 6), 7), 8), 9), 10), 11), 12), 13), 14), 15).
Per 1), 2) e pure 3), 5), 7), 13), 15), prima di determinare la moda converrà suddividere i dati in intervalli, o "classi" (esempio: le altezze in intervalli di 2-3 cm, il n° di amici su Facebook in intervalli di 100 …)

22) Una gara ciclistica per dilettanti in 3 tappe è stata vinta da un atleta che ha fatto registrare i tempi riportati in tabella. Quale è stata la velocità media dell'atleta nell'intera gara?

	km	*tempo*
Prima tappa	155	4h 42' 27"
Seconda tappa	94,5	2h 45' 08"
Terza tappa	147	4h 01' 45"

23) L'istogramma qui a destra (tracciato con OpenOffice Calc) → è relativo a un gruppo di ingegneri laureatisi a diverse età.

Quali sono la media, la mediana e la moda della distribuzione?

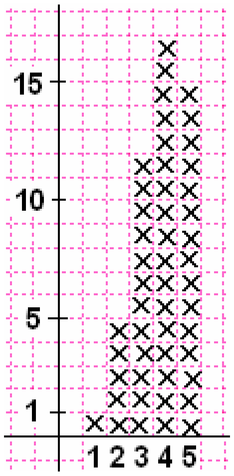


24) Con riferimento ai dati, già precedentemente considerati, e che qui sotto riportiamo, delle velocità di 60 auto controllate dalla Polizia, si domanda quali sono la media, la mediana e la moda della distribuzione.
Supponiamo ora di suddividere i dati in classi: da 45 km/h a 49 estremi inclusi, da 50 a 54, …
Quali sarebbero la *moda* (= *classe modale*), la *classe mediana* e la *media per classi* in questo caso?

57	61	52	48	68	48	55	56	72	49	50	55
54	61	58	63	64	61	47	52	53	59	48	54
56	60	51	51	50	78	67	61	58	55	59	53
62	54	49	45	56	60	51	52	60	54	55	51
48	57	56	55	58	53	59	70	74	64	81	52

Traccia (col foglio elettronico) l'istogramma dei dati suddivisi in classi.

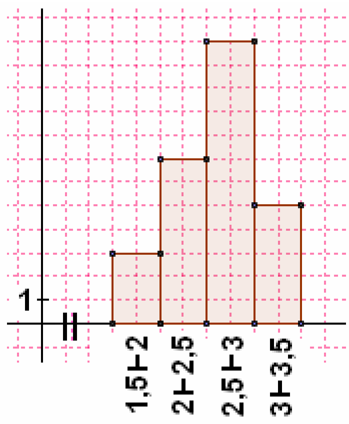
25) Cosa ti aspetti facendo la media degli esiti di tanti lanci di un dado? Lanciane effettivamente uno, almeno 50-100 volte (può essere un "lavoro di gruppo" … fossero tutti così, i lavori … ☺)



26) A un gruppo di residenti in un piccolo paese è stato chiesto di esprimere con un punteggio da 1 a 5 il proprio gradimento per la giunta comunale.

Le risposte si sono ripartite come illustrato dall'istogramma qui a sinistra.

Determina media, mediana e moda della distribuzione.



27) Andiamo a riprendere i dati registrati dall'insegnante di Educazione Fisica riguardo alle distanze saltate in lungo dai giovani allievi. Quali sono la media, la mediana e la moda della distribuzione?

28) La tabella sottostante, tratta da *Regards sur l'éducation 2008: Les indicateurs de l'OCDE* e relativa però a dati del 2006, mostra un indicatore della preparazione scientifica posseduta dagli studenti dei paesi aderenti all'organizzazione.
Con un foglio elettronico, ordina i dati e determinane la media e la mediana.
Fai poi comparire accanto a ciascun dato il suo scarto (positivo o negativo) dalla media.
Calcola la somma di questi scarti: cosa ti aspetti che esca?

Australie	527	Allemagne	516	Luxembourg	486	Espagne	488
Autriche	511	Grèce	473	Mexique	410	Suède	503
Belgique	510	Hongrie	504	Pays-Bas	525	Suisse	512
Canada	534	Islande	491	Nouvelle-Zélande	530	Turquie	424
Rép. chèque	513	Irlande	508	Norvège	487	Royaume-Uni	515
Danemark	496	Italie	475	Pologne	498	États-Unis	489
Finlande	563	Japon	531	Portugal	474		
France	495	Corée	522	Rép. slovaque	488		

29) Una pasticceria domanda alle famiglie dei suoi 3 dipendenti di assaggiare una nuova torta assegnandole un giudizio di gradimento da 0 a 5. Si decide però di attribuire peso 3 ai giudizi delle mamme, 2 a quelli dei figli, 1 a quelli dei papà. Se gli assaggiatori si sono espressi come segue, qual è la media finale?

M	P	F	F	M	P	F	F	F	M	P	F	F
4	4	5	4	3	3	4	4	4	3	4	5	4

30) Famiglie residenti in Italia classificate per numero di componenti
(valori assoluti in migliaia e composizioni percentuali) - Dati ISTAT

Numero di componenti	1961	1971	1981	1991	2001
1	10,6	12,9	17,9	20,6	24,9
2	19,6	22,0	23,6	24,7	27,1
3	22,4	22,4	22,1	22,2	21,6
4	20,4	21,2	21,5	21,2	19,0
5	12,6	11,8	9,5	7,9	5,8
6 o più	14,4	9,7	5,4	3,4	1,7
Totale	13747	15981	18632	19909	21811

Si può calcolare la media dei componenti di una famiglia in un dato anno, poniamo nel 1961?
E il numero approssimativo totale dei residenti in un dato anno, poniamo il 2001?

E) ALTRI TIPI DI MEDIA, OLTRE A QUELLA ARITMETICA

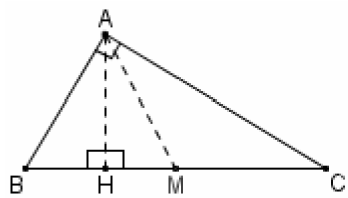
31) Per i seguenti dati determina
I) media aritmetica (senza usare né il computer né la calcolatrice)
II) media geometrica (calcolatrice: estrarre, ad es., la radice quinta, è come elevare all'esponente $1/5=0,2$)
III) media armonica (col computer: foglio elettronico)
IV) media quadratica (col computer: foglio elettronico)

a)	7	5	1	3	4
b)	1	1	1	2	1
c)	1	$1/2$	$1/4$		

32) Una scatola a forma di parallelepipedo rettangolo ha dimensioni (in cm) $30 \text{ X } 40 \text{ X } 50$.
Che spigolo dovrebbe avere un cubo (quindi: un parallelepipedo rettangolo con le 3 dimensioni uguali), se si desidera che il suo volume sia uguale a quello della scatola?
A quale quantità da noi studiata corrisponde la lunghezza dello spigolo di questo cubo?

33) Il "**quesito di Briatore**"
Tempo fa, a un noto VIP italiano, facente parte del mondo della Formula 1, venne posto il seguente quesito:
Qual è la velocità media di un'automobile in un circuito,
se metà dei giri sono coperti a 100 km/h e l'altra metà a 300 Km/h?
La risposta di Briatore fu (sorprendentemente, per alcuni) corretta.
Qual è questa risposta esatta? E come è presumibile che ci sia arrivato il VIP?

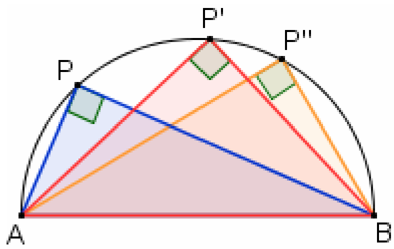
34) Considera, in un triangolo ABC rettangolo in A, l'altezza AH relativa all'ipotenusa e le due proiezioni BH e HC dei cateti sull'ipotenusa.
Il II° Teorema di Euclide afferma che vale la proporzione
$BH : AH = AH : HC$
Ma da ciò segue allora che AH rappresenta la media ………………
dei due segmenti BH e HC.



E a ben guardare, anche la mediana AM relativa all'ipotenusa può essere considerata come una media in relazione a BH e HC!
E' noto infatti che la mediana relativa all'ipotenusa in un triangolo rettangolo è metà dell'ipotenusa stessa; e ciò significa che la mediana AM rappresenta, dei due segmenti BH e HC, la media ……………………

35) L'esercizio 34) può servire a dimostrare geometricamente che dati due numeri positivi, la loro media geometrica non può mai essere maggiore della loro media aritmetica $(M_G \le M)$. Perché?

36) Si può dimostrare che se un angolo è "inscritto in una semicirconferenza" (vedi figura), allora è di 90°. $\rightarrow$
Perciò i triangoli ABP, ABP', ABP'', ... in figura sono tutti rettangoli;
Bene, le coppie di cateti hanno "qualcosa" in comune che ha a che fare con il discorso "medie".
Che cosa?



37) Percorro in motorino l'anello di 2500 metri intorno al mio isolato, tenendo il tachimetro sui 30 km/h al primo giro, sui 35 km/h al secondo e sui 45 km/h al terzo.
Qual è la mia velocità media sui tre giri?

38) Percorro in motorino l'anello che circonda il mio isolato, tenendo il tachimetro sui 30 km/h per 5 minuti, sui 35 km/h per altri 5' e sui 45 km/h per ulteriori 5'.
Qual è la mia velocità media in questo quarto d'ora?

39) Trova la velocità media nei seguenti due casi:
a) Si procede per ½ ora a 1 km all'ora, per ½ ora a 2 km all'ora e infine per un'altra ½ ora a 6 km all'ora
b) Si procede per ½ km a 1 km all'ora, per ½ km a 2 km all'ora e infine per un'altro ½ km a 6 km all'ora

40) Sono un ciclista dilettante, e mi alleno. Ho pedalato ¾ d'ora ai 24 km/h.
A che velocità dovrei procedere i successivi ¾ d'ora, se desiderassi ottenere una velocità media complessiva di 27 km/h?

41) Sono un ciclista dilettante, e mi alleno. Ho pedalato 1 quarto d'ora ai 24 km/h.
A che velocità dovrei procedere i successivi ¾ d'ora, se desiderassi ottenere una velocità media complessiva di 27 km/h?

42) Sono un ciclista dilettante, e mi alleno. Ho percorso 6 km ai 24 km/h. A che velocità dovrei coprire i 6 km restanti, se desiderassi ottenere una velocità media complessiva di 27 km/h?

43) Sono un ciclista dilettante, e mi alleno. Ho percorso 6 km in 1 quarto d'ora. A che velocità dovrei coprire i successivi 24 km del tragitto, se desiderassi ottenere una velocità media complessiva di 27 km/h?

44) Un "amico" mi ha persuaso a un investimento col quale ho guadagnato il 3% il 1° anno, ho guadagnato ancora il 5% il 2° anno (NOTA), e ho perso però poi l'8% il 3° anno ☹.
Qual è la mia situazione finanziaria dopo tutto ciò?

NOTA - Quando si dice "guadagno il p%", occorrerebbe sempre specificare *rispetto a che cosa* quel p% deve essere calcolato. In casi come il nostro, quando ci si riferisce a guadagni o perdite *anno dopo anno*, si intende che il p% sia da calcolarsi rispetto alla cifra che si possedeva *all'inizio dell'anno in questione.*

45) Con un foglio elettronico, traccia una “serie storica” che illustri l’evolversi di un capitale di 100 euro, i cui incrementi annui, in un triennio, siano stati rispettivamente del 3%, del 5% e del -8%.

46) Se un usuraio, dopo aver prestato 100, richiede 150 dopo 2 anni, è come se avesse applicato il tasso di interesse medio annuo del … ?

47) Nel giro di 2 anni, per via della crisi di una grande azienda, il valore delle sue azioni è dimezzato. Quale è stata la diminuzione percentuale media annua?
(Suggerimento: se ogni anno la diminuzione in percentuale fosse sempre stata la medesima, allora, indicando con x questa percentuale, dopo 1 anno il prezzo iniziale p ce lo saremmo ritrovato moltiplicato per $1-\frac{x}{100}$, dopo due anni per $\left(1-\frac{x}{100}\right)^2$, da cui l’equazione …)

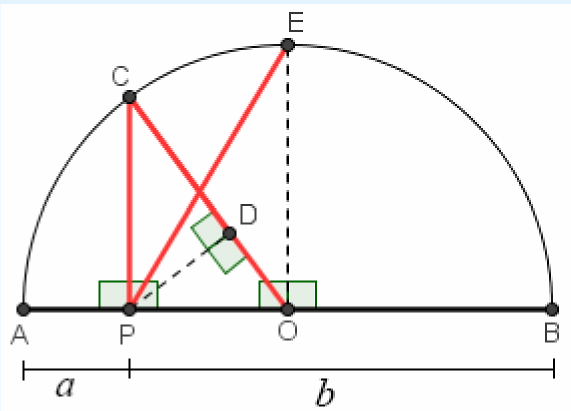
48) Un’azienda meccanica utilizza una vecchia apparecchiatura M_1 in grado di produrre 24 pezzi all’ora. Questa macchina viene lasciata in funzione per un tempo t_1, fino a che ha prodotto k pezzi. Successivamente viene spenta e al suo posto ne viene sperimentata un’altra, M_2, di ultima generazione, che lavora al ritmo di 40 pezzi all’ora. Questa seconda macchina viene lasciata in funzione per un tempo t_2, fino a che ha prodotto anch’essa k pezzi. Quanti pezzi all’ora dovrebbe produrre una 3^a macchina M_3, se si desidera che possa fabbricare k pezzi nella media aritmetica $\frac{t_1+t_2}{2}$ dei due tempi t_1 e t_2?
La risposta sta in una delle medie da noi studiate?

E terminiamo con due esercizi davvero molto belli, ma difficili.
Essi richiedono qualche nozione di Geometria che nelle scuole superiori italiane dovrebbe senz’altro essere acquisita entro il primo biennio (angoli inscritti in semicirconferenze, teoremi di Euclide o anche solo conoscenza delle Similitudini, ecc.).
Dimostra i seguenti enunciati:

49)

Sia $AP = a$, $PB = b$. Tracciamo la semicirconferenza di diametro AB, poi per P la perpendicolare al diametro fino a raggiungere la semicirconferenza in C, quindi il raggio OC, la distanza PD di P da OC, la perpendicolare per il centro O al diametro fino a raggiungere la semicirconferenza in E, la congiungente PE.
Allora i segmenti OC, PC, DC, PE sono altrettante medie fra a e b:

OC = media *aritmetica*;
PC = media *geometrica*
DC = media *armonica*;
PE = media *quadratica*



E da tutto ciò si può trarre che è sempre (per a, b positivi):

$$M_Q \ge M \ge M_G \ge M_A$$

50)

In un trapezio le due basi misurano a, b.

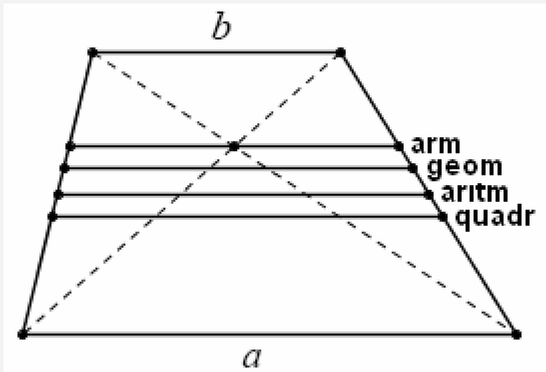
Allora quattro segmenti, ciascuno interno al trapezio e parallelo alle sue due basi, rappresentano altrettanti tipi di media fra a e b.

I) Il segmento equidistante dalle due basi ne rappresenta la media *aritmetica*

II) Il segmento, che ha la proprietà di dividere il trapezio in due trapezi simili fra loro, ne rappresenta la media *geometrica*

III) Il segmento, che ha la proprietà di dividere il trapezio in due trapezi aventi ugual area, ne rappresenta la media *quadratica*

IV) Il segmento che passa per il punto di intersezione delle due diagonali ne rappresenta la media *armonica*



Sugli INDICI DI DISPERSIONE (risposte a pag. 425)

Consider a **population** consisting of the following eight values: 2, 4, 4, 4, 5, 5, 7, 9

The eight data points have a **mean** (or **average**) value of 5: $\frac{1}{8}(2+4+4+4+5+5+7+9)=5$

To calculate the population standard deviation,
first compute the difference of each data point from the mean, and square the result:

$$(2-5)^2=(-3)^2=9 \quad (4-5)^2=(-1)^2=1 \quad (4-5)^2=(-1)^2=1 \quad (4-5)^2=(-1)^2=1$$
$$(5-5)^2=0^2=0 \quad (5-5)^2=0^2=0 \quad (7-5)^2=2^2=4 \quad (9-5)^2=4^2=16$$

Next divide the sum of these values by the number of values and take the square root to give the standard deviation: $\sqrt{\frac{1}{8}(9+1+1+1+0+0+4+16)}=2$

Da Wikipedia l'enciclopedia libera

1) Per ciascuna delle tre serie di dati riportati in tabella determina:
   I) campo di variabilità
   II) scarto assoluto medio
   III) scarto quadratico medio o deviazione standard,
   sia con la formula-base

$$\boxed{\sigma=\sqrt{\frac{(x_1-M)^2+(x_2-M)^2+...+(x_n-M)^2}{n}}}$$

   che con la formula alternativa $\boxed{\sigma=\sqrt{\frac{x_1^2+x_2^2+...+x_n^2}{n}-M^2}}$

   IV) coefficiente di variazione

a)	7	5	1	3	4
b)	1	1	1	2	1
c)	1	$\frac{1}{2}$	$\frac{1}{4}$		

2) Fra le locuzioni “scarto medio”, “scarto assoluto medio” e “scarto medio assoluto”, quale è più corretta?

3) In un campione di persone appartenenti a una riserva indiana viene misurata la lunghezza della spanna, e si trova una media di 18,0 cm con una deviazione standard di 0,8 cm mentre misurando la lunghezza del piede si trova una media di 24,2 cm e una deviazione standard di cm 1,1.
Il grado di “dispersione” delle misure è maggiore per la mano o per il piede?

4) Lavoro di gruppo in laboratorio.

Si raccolgono, in un foglio elettronico,
i numeri di scarpe di tutte le femmine della classe;
si fa lo stesso per tutti i maschi.

Poi si fa in modo che il foglio elettronico calcoli,
per ciascuno dei due insiemi di dati:
- la media;
- il campo di variabilità;
- lo scarto quadratico medio;
- il coefficiente di variazione.

Cosa dovrebbe guardare chi desiderasse farsi un’idea se siano più omogenee
le “dimensioni” della parte femminile oppure di quella maschile della classe?

5) La tabella che segue registra il numero di anni di permanenza al trono dei regnanti inglesi
da King Athelstan (924-940) a Queen Elizabeth II (in carica il 30/6/2010, data dell’ultimo aggiornamento).
Le durate sono state disposte in ordine crescente, e i dati arrotondati all’intero più vicino.

1	1	1	2	2	3	4	4	4	5	6	6	7	9	9	9	10	12	13	13	13	13
15	16	16	18	19	19	19	20	21	21	22	23	24	24	24	25	26	33	35	35	35	38
38	39	44	50	56	58	60	64														

a) Calcola, con un foglio elettronico, la media e lo scarto quadratico medio di questi dati.

b) Successivamente, raggruppa i dati in classi di 3 anni (da 1 a 3; da 4 a 6 …),
determina la frequenza di ciascuna classe e calcola nuovamente la media,
prendendo come valore corrispondente a ogni classe il valore centrale di quest’ultima.
Infine, fai lo stesso per classi di 5 anni (da 1 a 5, da 6 a 10 …).
Le medie così calcolate sono prossime fra loro?

6) Considera i seguenti tre insiemi di dati: A) 7 8 6 6 8 B) 9 7 7 5 7 C) 7 8 7 6 7
Si tratta dei voti di matematica presi da tre diversi studenti nel corso dell'anno scolastico.
Potrai verificare che la media è la medesima;
tuttavia, uno di questi ragazzi è stato, per così dire, più "costante" degli altri nelle sue *performance*.
Di chi si sta parlando: dello studente A, B o C?

7) Quattro padri e quattro figli, appassionati di atletica,
hanno fatto registrare i tempi seguenti in una corsa campestre:
25', 34', 27' e 29' i padri; 36', 45', 47' e 43' i figli.
Nel complesso, hanno avuto prestazioni più disomogenee i padri o i figli?

8) Una *software house* dispone di due uffici tecnici, uno a Torino e l'altro a Milano,
cui si rivolgono, tramite Internet, i clienti di tutta Italia per consulenze
riguardo a problemi vari inerenti alla installazione e gestione del software.
E' interessante analizzare l'insieme di dati costituito dal numero di minuti che sono intercorsi,
nei due uffici, fra l'apertura della email che richiedeva l'intervento e la prima email di risposta
finalizzata alla risoluzione del problema, per un campione di 50 interventi consecutivi.

Ufficio torinese:

10	38	36	10	12	14	45	43	42	16	41	37	13	6	11	39	48	39	44	2	9	15	7	14	40
41	8	10	13	12	42	8	10	38	36	41	39	39	16	8	39	39	42	45	43	7	9	39	44	16

Ufficio milanese:

68	72	68	68	65	25	29	75	30	32	36	17	24	77	32	36	64	67	75	21	75	29	73	67	69
28	76	39	33	38	75	37	37	30	80	32	70	28	25	67	70	30	31	72	71	67	78	67	64	29

Che ipotesi si possono fare sul modo di lavorare delle due squadre di tecnici?

9) La tabella sottostante, tratta da *Regards sur l'éducation 2008: les indicateurs de l'OCDE*
e relativa però a dati del 2006, mostra un indicatore della preparazione scientifica
posseduta dagli studenti dei paesi aderenti all'organizzazione.

Con un foglio elettronico, determina il campo di variabilità di questi dati
sulle competenze scientifiche degli studenti, senza ordinare i dati medesimi.
Determina pure scarto assoluto medio, varianza, scarto quadratico medio, e coefficiente di variazione.

Calcola altresì lo scarto quadratico medio "corretto", quello
che si ottiene con la formula a fianco e che, per n grande,
è molto prossimo allo scarto quadratico medio "non corretto".

$$\boxed{\sqrt{\frac{(x_1-M)^2+(x_2-M)^2+\ldots+(x_n-M)^2}{n-1}}}$$

Australie	527	Allemagne	516	Luxembourg	486	Espagne	488
Autriche	511	Grèce	473	Mexique	410	Suède	503
Belgique	510	Hongrie	504	Pays-Bas	525	Suisse	512
Canada	534	Islande	491	Nouvelle-Zélande	530	Turquie	424
Rép. chèque	513	Irlande	508	Norvège	487	Royaume-Uni	515
Danemark	496	Italie	475	Pologne	498	États-Unis	489
Finlande	563	Japon	531	Portugal	474		
France	495	Corée	522	Rép. slovaque	488		

10) E sai a cosa si riferiscono questi dati, tratti dalla stessa fonte di prima?
Al numero medio di ore annue obbligatorie passate a scuola da uno studente di 15 anni (nel 2006).

Australie	Autriche	R. tch.	Danemark	Angleterre	Finlande	France	Allemagne	Grèce
968	1005	960	900	760	856	1033	900	1117
Hongrie	Islande	Irlande	Italie	Norvège	Portugal	Espagne	Suède	Turquie
763	888	802	1089	855	826	979	741	810

Richieste come per l'esercizio precedente.

11) Dimostra che $\sigma=\sqrt{\frac{(x_1-M)^2+(x_2-M)^2+\ldots+(x_n-M)^2}{n}}=\sqrt{\frac{x_1^2+x_2^2+\ldots+x_n^2}{n}-M^2}$, nel caso $n=2$.

12) La quantità $\sqrt{\frac{(x_1-a)^2+(x_2-a)^2+\ldots+(x_n-a)^2}{n}}$ è minima quando $a=M=\frac{x_1+x_2+\ldots+x_n}{n}$.
Con un foglio elettronico, verifica questo fatto su di un esempio.

11. GLI ERRORI DI MISURA

Prendi un metro da muratore (di quelli pieghevoli, formati da più aste collegate da cerniere, totale 2 m) e prova a misurare, al centimetro, la lunghezza del corridoio della tua scuola.
Ripeti l'operazione più volte, segnando sempre su di un taccuino il valore ottenuto.

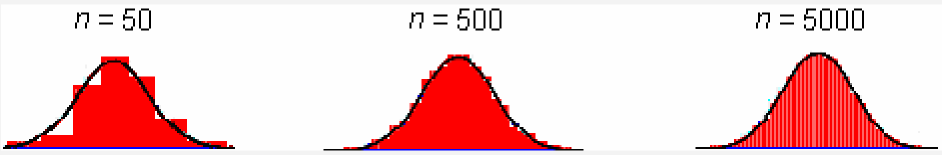
Certamente non otterrai la stessa misura ad ogni prova:
infatti, nel disporre il metro sul pavimento, ti capiterà di non iniziare esattamente dallo stesso punto, di riportare il metro non sempre con precisione quando devi spostarlo per ricollocarne un'estremità nella posizione alla quale eri giunto al passo precedente, di piegarlo leggermente, e così via.

Adesso coraggio, perché ho bisogno che tu faccia TANTE misurazioni, diciamo 100
(sono certo che i tuoi compagni di classe si presteranno a collaborare … ognuno potrebbe fare 4-5 misurazioni).

Ora hai a disposizione 100 numeri.
Può darsi che alcuni di questi numeri coincidano, ma in generale saranno invece un poco diversi fra loro.
Considera il minimo e il massimo valore rilevato, e suddividi l'intervallo $[x_{\min}, x_{\text{MAX}}]$ in un certo numero di sottointervalli, diciamo otto-dieci
(in generale, se le misure sono *n*, si consiglia di far sì che il numero di intervalli non superi $\sqrt{n}$) :
ad esempio, se la minima e la massima delle misure registrate sono state di m 23,92 e di m 24,11, avremo $24{,}11-23{,}92=0{,}19$ e questo intervallo di metri 0,19 (19 cm) potrà portarci a definire 10 sottointervalli di 2 cm ciascuno: $[23{,}92;\, 23{,}94),\ [23{,}94;\, 23{,}96),\ [23{,}96;\, 23{,}98),\ \ldots,\ [24{,}10;\, 24{,}12)$.
Ora, per ciascun sottointervallo, conta la rispettiva "frequenza",
ossia conta il numero di misure, fra le 100 registrate, che cadono in quel sottointervallo;
traccia, con un foglio elettronico, un istogramma con le *classi di misura* in *orizzontale* e le *frequenze* in *verticale*.

Potrai osservare che le misure "centrali" della distribuzione saranno in linea di massima più frequenti, e quelle estreme meno. In effetti, nell'atto pratico della misurazione, si commettono sempre **errori "casuali"** talvolta in difetto talvolta in eccesso, e **se il numero di misurazioni effettuate diventa alto, l'istogramma tenderà ad assomigliare a una curva "a campana"** detta "**gaussiana**" (F. Gauss, 1777-1855).

Ecco qui di seguito un "fumetto" di possibili configurazioni dell'istogramma delle frequenze al crescere del numero n di misure effettuate.



La **Gaussiana** è una curva la cui equazione è nientemeno che $y=\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, dove:

$\pi=3{,}14159...$ (ben noto); $e=2{,}71828...$ (numero di Nepéro);
$\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ sono due numeri fissi che, nel caso in cui la curva abbia a che fare con il problema da noi esaminato, ossia quello delle misure ripetute di una quantità (affette da errori "casuali" o "statistici"), sono interpretabili come rispettivamente la **media aritmetica** e lo **scarto quadratico medio che si otterrebbero facendo un numero colossale (= *tendente all'infinito*) di misure**.

Trovi la cosa complicata? In effetti, lo è … ☹
Questi studi richiedono nozioni matematiche più avanzate (la teoria delle "*distribuzioni di probabilità*")
e non è facile, in una trattazione di carattere non specialistico,
mantenere il discorso su di un livello che sia nel contempo accessibile e rigoroso …

… ma noi ci proviamo ☺.

**Se le misurazioni effettuate, affette da errore casuale, sono tante
(di solito, detto *n* il numero di misure, "tante" significa perlomeno $n>30$,
ma alcuni Autori scrivono $n>50$ o $n>60$, altri $n>100$ … e insomma, più sono, meglio è),
allora l'istogramma delle frequenze tende ad assomigliare ad una gaussiana**;
e *quanto più* tale somiglianza sussiste, *tanto più,*

detta $\bar{x}$ la media di queste misure $\boxed{\bar{x}=\textbf{MEDIA}=\frac{x_1+x_2+\ldots+x_n}{n}}$

e detto s il loro scarto quadratico medio $\boxed{s=\textbf{S.Q.M.}=\sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_n-\bar{x})^2}{n}}}$,

sono corrette le affermazioni seguenti:

a) **$\overline{x}$ è un valore prossimo al vero valore della grandezza in questione, dove per "vero" valore si intende quello che si otterrebbe come media su di un numero enorme di misure**

b) **circa il 68% delle misure effettuate rientra nell'intervallo $(\overline{x}-s,\ \overline{x}+s)$**
circa il 95% delle misure effettuate rientra nell'intervallo $(\overline{x}-2s,\ \overline{x}+2s)$
circa il 99,7% delle misure effettuate rientra nell'intervallo $(\overline{x}-3s,\ \overline{x}+3s)$

c) **se facessi un'ulteriore misura, questa avrebbe**
circa il 68% di probabilità di cadere nell'intervallo $(\overline{x}-s,\ \overline{x}+s)$
circa il 95% di probabilità di cadere nell'intervallo $(\overline{x}-2s,\ \overline{x}+2s)$
circa il 99,7% di probabilità di cadere nell'intervallo $(\overline{x}-3s,\ \overline{x}+3s)$

Di solito, per misurare una grandezza fisica, si effettua un certo numero *n* di operazioni di misura, si calcolano la media $\overline{x}$ e lo scarto quadratico medio s degli *n* valori così trovati, poi si scrive che la grandezza in gioco vale

$$\boxed{\overline{x} \pm s},$$

dove per la piena comprensione di questa scrittura occorre tenere presenti le 3 considerazioni a), b), c).

Torniamo soltanto a ribadire alcuni concetti davvero fondamentali.
Affinché le affermazioni precedenti siano corrette, il numero *n* delle misure deve essere "GRANDE"…

Inoltre LE AFFERMAZIONI CONTENGONO DEGLI AVVERBI "CIRCA", NON SOLO PER IL FATTO CHE I VALORI 68%, 95%, 99,7% SONO TUTTI APPROSSIMATI, MA SOPRATTUTTO PER IL FATTO CHE SI STA PENSANDO AD UNA CONFIGURAZIONE PROBABILISTICA IDEALE ALLA QUALE SI TENDE AD AVVICINARSI (SENZA PERO' RAGGIUNGERLA) AL CRESCERE DI *n* .
La veridicità di a), b), c) è tanto maggiore quanto più $\overline{x}$ (media delle n misure realmente effettuate) è prossimo a μ (vero valore della grandezza, media su un numero di misure che tende all'infinito) e quanto più *s* (s. q. m. delle n misure realmente effettuate) è prossimo a σ (s. q. m. su "infinite" misure); ed è all'aumentare del numero delle misure che effettivamente $\overline{x}, s$ tendono a identificarsi con μ, σ !!!

ESEMPIO

Qui sotto riportiamo 96 misure in mm della larghezza della lavagna di un'aula,
rilevate dai 24 studenti, che hanno effettuato 4 misurazioni ciascuno:

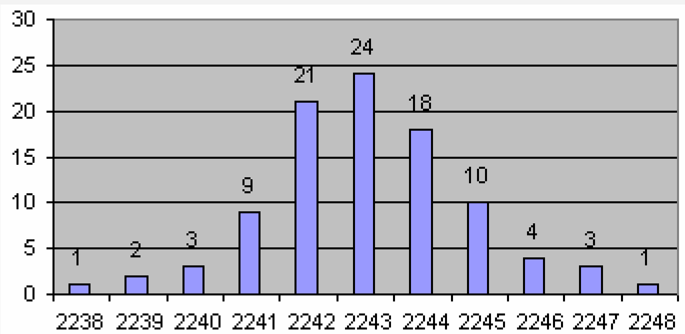
2242	2240	2244	2243	2244	2244	2242	2247	2244	2242	2246	2244	2241	2244	2242	2241	2244	2243	2242	2241
2242	2242	2243	2242	2246	2243	2245	2238	2246	2244	2244	2244	2243	2242	2245	2241	2243	2239	2244	2243
2245	2243	2247	2243	2244	2245	2242	2243	2245	2239	2246	2242	2243	2241	2244	2245	2244	2241	2242	2241
2243	2242	2243	2244	2243	2248	2243	2242	2241	2245	2243	2242	2240	2245	2244	2242	2243	2242	2241	2243
2243	2244	2243	2242	2245	2244	2243	2242	2243	2245	2242	2240	2243	2242	2243	2247				

Il calcolo ci dà

$$media = \overline{x} = 2243{,}06...; \quad scarto\ quadratico\ medio = s = 1{,}79...$$

Se ora andiamo a contare il numero di misure che sono comprese nell'intervallo $(\overline{x}-s,\ \overline{x}+s)$, vediamo che tali misure sono $21+24+18=63$.
Bene, 63 è assai prossimo al 68% di 96 (che vale circa 65). [vedi NOTA]

Ecco l'istogramma della distribuzione di frequenza, che in effetti presenta, pur con irregolarità, il tipico andamento "a campana".



NOTA - Per la precisione, quello che abbiamo inizialmente indicato come il 68% avrebbe potuto essere meglio approssimato come 68,3%, e il 95% come 95,4%. Oppure, si sarebbe potuto scrivere 95% ma sostituendo il fattore 2 con un più preciso 1,96.

Lo diciamo per scrupolo, e tuttavia insistiamo: **non dobbiamo confondere la configurazione probabilistica ideale, teorica, alla quale ci si avvicinerebbe se *n* tendesse a infinito, con la situazione reale, che è approssimata bene, ma non certo alla perfezione, quando *n* comincia ad esser >30, o meglio ancora >100.**

Un'ultima puntualizzazione.
La Statistica Inferenziale insegna che, per meglio stimare lo scarto quadratico medio σ relativo alle "infinite" misure, è più giusto calcolare lo scarto quadratico medio s del "campione di n misure" attraverso la formula "corretta" che si ottiene prendendo come denominatore $n-1$ anziché n:

$$\text{scarto quadratico medio "corretto"} = \hat{s} = \sqrt{\frac{\left(x_1-\overline{x}\right)^2+\left(x_2-\overline{x}\right)^2+\ldots+\left(x_n-\overline{x}\right)^2}{n-1}} = \sqrt{\frac{n}{n-1}}\cdot s$$

E' pur vero che **quando n è grande, scarto quadratico medio "corretto" e "non corretto" differiscono di pochissimo; e quando n non è grande, la teoria esposta non vale più!**
Già per valori di n dell'ordine di qualche decina, la differenza è assai piccola.
Ad esempio, con $n=30$, il fattore $\sqrt{n/(n-1)}$ vale $\sqrt{30/29}\approx 1{,}017$ che è molto vicino a 1!

In **EXCEL** e in **OPENOFFICE CALC**
lo **scarto quadratico medio "non corretto**" è **dev.st.pop()** mentre quello "**corretto**" è **dev.st()**

APPROFONDIMENTO (NON SEMPLICE): INTERVALLI DI CONFIDENZA, ERRORE STANDARD
In realtà, quando andiamo a calcolare la media $\overline{x}$ e lo scarto quadratico medio s sulle n misure che abbiamo effettuato, il nostro interesse è puntato, più che a *quelle* particolari n misure, al valore "*vero*" – che ci è sconosciuto – della grandezza in esame.
Ora, abbiamo già detto che quest'ultimo può essere pensato come "quel valore μ che si otterrebbe come media su di un numero sterminato di misure".
Ma fino a che punto possiamo ritenere che la media $\overline{x}$ da noi calcolata sia prossima al "vero" valore μ?
La "statistica inferenziale" ci insegna che se noi effettuiamo una serie di n misure, ed n è grande (certi Autori scrivono $n>30$, altri $n>50$ o 60, altri ancora $n>100$; … in realtà … quanto stiamo dicendo tende ad essere tanto più veritiero quanto più n è alto), allora, determinando per queste n misure la media $\overline{x}$ e lo scarto quadratico medio s, il *vero* valore μ avrà una probabilità

- **del 68% circa di rientrare nell'intervallo** $\left(\overline{x}-\frac{s}{\sqrt{n}},\ \overline{x}+\frac{s}{\sqrt{n}}\right)$
- **del 95% circa di rientrare nell'intervallo** $\left(\overline{x}-2\frac{s}{\sqrt{n}},\ \overline{x}+2\frac{s}{\sqrt{n}}\right)$
- **del 99,7% circa di rientrare nell'intervallo** $\left(\overline{x}-3\frac{s}{\sqrt{n}},\ \overline{x}+3\frac{s}{\sqrt{n}}\right)$

anche se per *maggiore precisione concettuale*, poiché il "vero valore" è … quello che è, è *costante*, mentre *a variare è invece l'insieme delle n misure e con esso l'intervallo* che ne deriva (è come se noi "estraessimo a sorte un intervallo, per poi domandarci se comprende o no il valore 'vero' "), bisognerebbe piuttosto partire "dal punto di vista dell'intervallo", dicendo che

- **il 68% circa degli intervalli** $\left(\overline{x}-\frac{s}{\sqrt{n}},\ \overline{x}+\frac{s}{\sqrt{n}}\right)$
  costruiti ciascuno facendo n misure e calcolandone i relativi $\overline{x}$ e s
  contiene al suo interno il "vero" valore (e il 32% circa lo lascia invece al suo esterno)
- **il 95% circa degli intervalli ecc. ecc.**
- **il 99,7% circa degli intervalli ecc. ecc.**

Questi intervalli di cui abbiamo parlato vengono chiamati "INTERVALLI DI CONFIDENZA".
Ad es., $\left(\overline{x}-3\frac{s}{\sqrt{n}},\ \overline{x}+3\frac{s}{\sqrt{n}}\right)$ **è un "intervallo di confidenza al 99,7%" per il vero valore della grandezza.**

Osserviamo l'uso del termine "confidenza" (= fiducia) al posto di "probabilità".

La quantità $\frac{\sigma}{\sqrt{n}}$ viene detta "**ERRORE STANDARD DELLA MEDIA**" (brevemente: "errore standard"), e, se n è grande, così come $\overline{x}$ è una buona approssimazione per μ, allo stesso modo s è una buona approssimazione per σ e quindi $\frac{s}{\sqrt{n}}$ è una buona approssimazione per $\frac{\sigma}{\sqrt{n}}$.

Il discorso è intrigante, ma complicato. RICAPITOLIAMO LE PREMESSE E LA SIMBOLOGIA.

Stiamo supponendo di ricercare il valore "vero" di una determinata grandezza, tramite una misura, anzi: tramite una serie di n misure, di cui faremo poi la media.
μ è il vero valore della grandezza. μ è incognito e viene approssimato con la media $\overline{x}$ delle n misure realizzate.

Se noi avessimo la possibilità di effettuare un numero grandissimissimissimo di misure, al tendere all'infinito di questo numero, la media delle misure tenderebbe a μ.

Ma noi per forza di cose ci dobbiamo accontentare delle nostre n misure.
n è grande, ma non colossale: prenderemo n almeno maggiore di 30, preferibilmente maggiore di 100 …
tuttavia le nostre misure, pur essendo tante, saranno n e basta.

Calcoleremo dunque la media $\overline{x}$ e lo scarto quadratico medio s delle nostre n misure.
Bene, avendo preso n piuttosto grande abbiamo fiducia che $\overline{x}$ sia già una approssimazione piuttosto precisa per μ, e che s sia già prossimo a quello che sarebbe lo s. q. m. σ se noi potessimo effettuare "infinite" misure.

Possiamo anzi "quantificare" questa nostra "fiducia".

Se consideriamo, ad esempio, l'intervallo $\left(\overline{x}-2\frac{s}{\sqrt{n}},\ \overline{x}+2\frac{s}{\sqrt{n}}\right)$,

la nostra fiducia che questo intervallo contenga μ è all'incirca del 95%, perché la Statistica Inferenziale insegna che, qualora andassimo a barbosissimamente effettuare 100 serie, o 1000 serie, … , di n misure ciascuna, calcolando per ognuna di queste il relativo $\overline{x}$ e il relativo s,

all'incirca il 95% degli intervalli $\left(\overline{x}-2\frac{s}{\sqrt{n}},\ \overline{x}+2\frac{s}{\sqrt{n}}\right)$ così costruiti conterrebbero μ.

Questo è tanto più vicino al vero quanto più n è grande, ma a partire da $n>30$ cominciamo già ad andar benino!

ESEMPIO

Prendiamo in prestito un esempio dal testo "**Essential medical statistics**" di B. R. Kirkwood e J. A. C. Sterne, dove ogni cosa è spiegata con calma, precisione, e ottimi riferimenti concreti (*hats off*, tanto di cappello!)

In realtà qui si ragiona in un ambito più generale del nostro.
Viene infatti esaminata non una singola grandezza misurata più volte, bensì una "*popolazione*" limitata (l'insieme delle 10000 case), nonché un suo "*campione*" (le 100 case che vengono visitate).
Ma lo stesso discorso fatto per le misure vale, nei suoi tratti essenziali, anche in questo contesto, perché si può osservare che la quantità di cui ci si sta occupando (la superficie da disinfestare nelle case) è uno degli svariati fenomeni della realtà che presentano una *Gauss-like distribution*, vale a dire: una distribuzione simile alla gaussiana.

Nell'ambito di un piano per l'eradicazione della malaria si progetta di trattare con insetticida tutte le 10000 case di una certa area rurale.
Problema: quanto insetticida acquistare?
Per deciderlo, si estrae da quelle 10000 case un campione casuale di 100 case, e le si ispeziona per misurare in ciascuna casa la superficie che richiede di essere bonificata.

In quelle 100 case la superficie media su cui spruzzare l'insetticida risulta essere di $\overline{x}=24{,}2\ m^2$ *con uno scarto quadratico medio* $s=5{,}9\ m^2$.

Non è realistico a questo punto supporre che la superficie media $\overline{x}$ *rilevata nel campione di 100 case coincida con la media* μ *della superficie da disinfestare nell'intera "popolazione" delle 10000 case; tuttavia, è possibile valutare quanto sia da ritenere affidabile la media campionaria* $\overline{x}=24{,}2\ m^2$

se si va a calcolare l'errore standard, approssimabile con $\frac{s}{\sqrt{100}}=\frac{s}{10}=\frac{5{,}9}{10}=0{,}59\approx 0{,}6$.

A questo punto, infatti, si può dire che l'intervallo $(24{,}2\pm 0{,}6)\ m^2$ *ha una probabilità del 68% circa di contenere il valore incognito* μ *della media di tutta la "popolazione" delle 10000 case; e che l'intervallo* $(24{,}2\pm 2\cdot 0{,}6)\ m^2=(24{,}2\pm 1{,}2)\ m^2$ *ha una probabilità del 95% circa di contenere* μ.

Allora l'intervallo $(24{,}2\pm 1{,}2)\ m^2$ *è un intervallo di confidenza al 95% per* μ*; se quindi ipotizziamo che questo intervallo contenga* μ*, abbiamo una probabilità del 95% circa di ipotizzare il vero.*

μ *dovrebbe perciò, al 95% di "confidenza", di "fiducia", non essere superiore a* $(24{,}2+1{,}2)\ m^2=25{,}4\ m^2$ *per cui se acquistiamo una quantità di insetticida tale da poter coprire* $25{,}4\cdot 10000\ m^2=254000\ m^2$ *abbiamo il 95% di probabilità che questo sia sufficiente al bisogno.*

Tutto il discorso fatto regge bene perché la numerosità del nostro campione (n = 100) è decisamente alta.
Coraggio, allora: abbiamo stimato quanto insetticida plausibilmente ci serve, andiamo a procurarcelo.

E se volessimo comprare l'insetticida sulla base di una confidenza del 99,7% circa?
Per quanti metri quadrati dovremmo attrezzarci? Fai tu il semplice calcolo: troverai circa $260000\ m^2$.

MEAN ± SD oppure MEAN ± SEM ???

Il ruolo dello *scarto quadratico medio* (**SD, Standard Deviation**) e quello dell'*errore standard della media* o semplicemente *errore standard* (**SEM, Standard Error of the Mean**) non devono essere confusi.

Sovente alcuni risultati, ad esempio in Medicina, vengono scritti con un'incertezza uguale al SEM, che è sempre per definizione minore della SD, proprio per dare l'idea di una minore variabilità … ma ciò può essere fonte di fraintendimenti gravi, se il lettore poi confonde questo SEM con la SD.

Cerco di spiegarmi. Supponiamo che una certa caratteristica quantitativa x relativa al sangue umano venga testata su di un campione di 400 individui presi a caso dalla popolazione generale, e si trovi che in questi individui la caratteristica in gioco vale 235 ± 42, essendo 235 la media calcolata sui 400 individui osservati, e 42 la SD delle 400 osservazioni.

Supponiamo inoltre che si sappia che la caratteristica studiata si distribuisce nella popolazione secondo la "campana" di Gauss o comunque una sua buona approssimazione (NOTA ♥)

Bene, se si scrive che la caratteristica in esame è stata osservata, in quel campione di 400 soggetti, con un valore dato da 235 ± 42 $(mean,\ SD)$, allora un medico che legge l'articolo scientifico potrà dire:

> in quel campione di 400 persone, pressappoco il 95% aveva quel valore compreso fra $235 - 2 \cdot 42 = 151$ e $235 + 2 \cdot 42 = 319$, e siccome quel campione (essendo abbastanza numeroso) è un'immagine piuttosto fedele dell'intera popolazione, se si presenta da me un paziente che ha quel valore minore di 151 o maggiore di 319, sono portato a classificare quel caso come anomalo e tale da richiedere ulteriori indagini cliniche; se invece in un paziente il valore è esterno all'intervallo $235 \pm 2 \cdot 2{,}1$ (2,1 è il valore approssimato dell'Errore Standard della Media o SEM, che si desume da una SD di 42 con $n = 400$: $42/\sqrt{400} = 42/20 = 2{,}1$), questo non mi preoccuperà affatto!
>
> Piuttosto, l'intervallo $235 \pm 2 \cdot 2{,}1$ è un *intervallo di confidenza* al 95% per x, nel senso che ha il 95% di probabilità di contenere il "vero valore" di x, ossia la media dei valori di x nell'intera popolazione.

Quindi

- ♪ il **SEM** mi interessa per valutare con quale probabilità un dato intervallo intorno alla media campionaria contenga la media dell'intera popolazione, ossia per la **STIMA DELLA MEDIA INCOGNITA** μ,
- ♫ mentre la **SD** mi interessa per quantificare la **DISPERSIONE delle rilevazioni NEL MIO CAMPIONE**, considerazioni che poi posso estendere tali e quali all'intera popolazione, perché, dato il numero elevato di elementi del campione e dato che erano stati estratti casualmente dalla popolazione, **il campione rappresenterà abbastanza fedelmente la popolazione intera**.

NOTA IMPORTANTE ♥

Questa richiesta è essenziale, perché

PARECCHI FENOMENI DELLA REALTA' PRESENTANO UNA "GAUSS-LIKE DISTRIBUTION", MA CIÒ NON VALE PER ALTRI!

Ad esempio, hanno una distribuzione più o meno sovrapponibile alla gaussiana

- gli errori di misura, come abbiamo visto (ma, a dire il vero, *non proprio sempre*)
- le distanze dal centro di un bersaglio per una serie di tiri
- i quozienti di intelligenza
- le altezze degli adulti di una stessa etnìa e sesso

… mentre per la distribuzione dei *pesi* delle persone la differenza rispetto alla gaussiana è già più marcata.

SINONIMO di "**DISTRIBUZIONE GAUSSIANA**" è "**DISTRIBUZIONE NORMALE**".

Le considerazioni sopra riportate possono rendere una prima idea di alcune fra le questioni di cui si occupa la **STATISTICA INFERENZIALE**.

Essa **interviene quando si cerca di studiare una caratteristica dell'intera popolazione tramite osservazioni condotte su di un suo sottoinsieme ("campione"), e occorre quantificare il grado di attendibilità di questo procedimento**.

Come nei **sondaggi elettorali**.

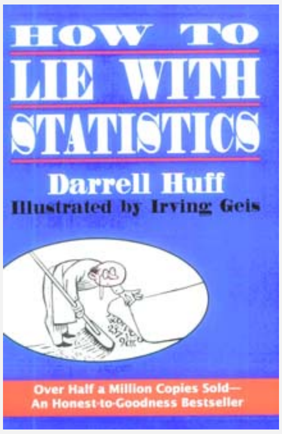
Come nelle **ricerche farmacologiche**, dove si va a confrontare l'evoluzione clinica di due gruppi di malati, a uno dei quali viene somministrata la sostanza attiva e all'altro, invece, un preparato inerte (il "*placebo*").

Come nei **test finalizzati a verificare** (in un determinato contesto) **la bontà di una ipotesi**.

La statistica inferenziale considera anche il caso in cui siano disponibili solo **piccoli campioni**.

Noi però, nei limiti del nostro corso, ci dobbiamo fermare ai pochi cenni dati, senza approfondire oltre.

How to Lie with Statistics



E' un libretto di divulgazione, scritto da Darrell Huff nel lontano 1954, che ha avuto uno straordinario successo di vendite, e conserva ancor oggi piena attualità.
Aiutandosi con garbate illustrazioni, passa in rassegna i modi attraverso i quali la pubblicità e la politica manipolano e presentano in modo parziale e distorto le statistiche, per spingere il consumatore o l'elettore a conclusioni sbagliate.

Il campione con l'errore incorporato
Alle interviste sulle letture abituali probabilmente la gente risponderà mentendo, almeno parzialmente, perché "confessare" letture frivole o imbarazzanti non fa fare bella figura. Analogo il discorso per l'igiene personale.
Poi le persone contattate hanno tendenza a dare quelle risposte che pensano possan far piacere a chi conduce l'intervista: l'autore riferisce, come esempio significativo, di un'analisi statistica a soggetto politico che aveva avuto esiti radicalmente diversi con intervistatori di pelle bianca o rispettivamente nera.
Le persone in una stazione ferroviaria sono rappresentative della popolazione generale? Probabilmente no: le madri di bambini piccoli, ad esempio, potrebbero scarseggiare in quel campione.
E gli incaricati a svolgere sondaggi per strada potranno tendere a scegliere persone più pulite o più gradevoli, o chi intuiscono sia più disponibile a rispondere, specie se devono terminare il loro compito in tempi ristretti.

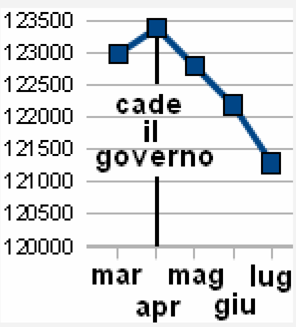
La media ben scelta
Quando si parla di "media", in realtà ci si sta riferendo a una *media aritmetica*, a una *mediana* o a una *moda*?
Dire che lo stipendio medio annuo dei dipendenti di un'azienda è, poniamo, di 38.500 dollari, può essere comodo per i dirigenti. Ma questa media, che è l'ordinaria *media aritmetica*, è comprensiva anche dei compensi stratosferici dei pochissimi manager strapagati, e i sindacalisti potrebbero invece considerare come stipendio "medio" la *mediana* degli stipendi, pari a 20.000 dollari (in pratica: solo metà dei dipendenti percepisce uno stipendio superiore a 20.000 dollari, l'altra metà inferiore). E' possibile che dirigenza e sindacati usino dunque il medesimo termine "media", in relazione a indicatori ben diversi.

Quei piccoli numeri che non ci sono
E' purtroppo frequente che si utilizzino (senza segnalarlo), per un'indagine statistica, campioni troppo piccoli per poter dare risultati attendibili; che si ometta la specificazione del grado di "dispersione" dei dati …

Molto rumore per praticamente nulla
Del tutto inutile confrontare due dati non molto differenti fra loro, senza specificare quale sia l'intervallo di "incertezza" di questi dati! Oppure: se abbiamo l'elenco completo delle marche di sigarette in commercio, elencate per grado di pericolosità decrescente ma tutte pressappoco allo stesso livello di tossicità, è insensato e ingannevole pubblicizzare la marca che sta in fondo all'elenco dicendo che è la "più raccomandabile"!

Il grafico fantasmagorico
"Tagliare" in modo scaltro i grafici, e/o scegliere furbamente le lunghezze dei segmenti che rappresentano date quantità in orizzontale e in verticale, può favorire forti distorsioni nella percezione di chi osserva.
Il diagramma in basso a sinistra nella pagina mira maliziosamente a suggerire che il numero di copie vendute da una rivista di politica sia crollato dopo la caduta del governo, mentre si è mantenuto pressoché stabile.

L'immagine monodimensionale
Rappresentare una quantità con ideogrammi ha un'insidia: se due raffigurazioni "in scala" di altezza una doppia dell'altra vengono utilizzate per illustrare il fatto che un certo valore è raddoppiato, l'osservatore ha comunque un'impressione diversa: l'area della seconda figura è *quadrupla* rispetto alla prima, e se anzi le figure vengono pensate come tridimensionali un raddoppio dell'altezza comporta un volume che è addirittura 8 volte tanto.
Quindi disegni di questo tipo possono essere impiegati per indurre la sensazione di una crescita (o diminuzione) più forte di quella reale. L'immagine in basso a destra dà un esempio di questo effetto psicologico.



Tracollo … o sostanziale stabilità delle vendite?

Il numero pseudoconnesso
"Se qualcuno non può dimostrare ciò che vorrebbe dimostrare, può dimostrare qualcos'altro e far finta che sia la stessa cosa" …

Il vecchio *post hoc* ritorna in sella
Se B segue, in ordine di tempo, A, ciò non implica che A sia causa di B.
Nelle Nuove Ebridi si era convinti che i pidocchi facessero bene alla salute ☺; in realtà, è ben facile che una persona ammalata sviluppi la febbre, e l'aumento di temperatura … scaccia i pidocchi! Causa ed effetto completamente ribaltati.

Il signore il basso guadagna… il doppio, il quadruplo, o 8 volte tanto?

ALTRI MODI DI QUANTIFICARE L'INCERTEZZA DELLA MISURA (per la distinzione fra la parola "errore" - spesso adoperata impropriamente - e la parola "incertezza", vedi l'importante NOTA a pag. 421)

a) **SCARTO ASSOLUTO MEDIO (SCARTO MEDIO, DEVIAZIONE MEDIA, ERRORE MEDIO)**
Al posto dello scarto quadratico medio s, si può prendere lo "scarto assoluto medio" δ
ossia la **media dei valori assoluti degli scarti dalla media**: $\boxed{\delta = \frac{|x_1-\bar{x}|+|x_2-\bar{x}|+\ldots+|x_n-\bar{x}|}{n}}$.

Si scriverà allora che la grandezza in esame vale $\boxed{\bar{x} \pm \delta}$

b) **SEMIDISPERSIONE (INCERTEZZA ASSOLUTA, ERRORE ASSOLUTO, ERR. MASSIMO)**
Effettuate le n misure $x_1, x_2, \ldots, x_n$ e calcolata la media $\bar{x}$ di queste, si va a determinare
la "semidispersione" (da alcuni detta "incertezza assoluta" o "errore assoluto" o "errore massimo")
cioè la **semidifferenza fra la più grande e la più piccola delle misure rilevate**: $\boxed{d = \frac{x_{\mathrm{MAX}} - x_{\mathrm{min}}}{2}}$

poi si scrive semplicemente che il valore della grandezza in questione è $\boxed{\bar{x} \pm d}$.

Questo metodo molto elementare della semidispersione viene impiegato più che altro QUANDO IL NUMERO DELLE MISURE A DISPOSIZIONE È BASSO O MOLTO BASSO.

La semidispersione è sovente indicata col simbolo Δx (naturalmente, se la grandezza è t, si userà Δt !)
Si legge "delta *x*"; quel Δ è un simbolo utilizzato, in questo e in altri casi, come "operatore di differenza".

c) **IL CASO DELLA MISURA UNICA, AD ES. PERCHE' LO STRUMENTO E' POCO SENSIBILE**
Quando, infine, lo strumento di misura è poco sensibile, cosicché
gli errori "casuali" o "statistici" non emergono e si rileva dunque sempre la stessa, grossolana, misura;
oppure anche quando l'operazione di misura viene effettuata una sola volta,
si scrive, detta $\bar{x}$ la misura trovata, che il valore della grandezza è $\boxed{\bar{x} \pm a}$,
essendo a l'**ampiezza** dell'**intervallo che corrisponde a due "tacche" consecutive del misuratore**
(o la **semiampiezza** nel caso le tacche siano abbastanza distanziate).

In qualsiasi caso, l'incertezza dichiarata riguardo a una misura non dovrebbe mai essere inferiore a quella dovuta alla sensibilità dello strumento.

"*In generale, la presenza di errori casuali nella misura fa sì che l'errore statistico risulti maggiore dell'errore strumentale (la sensibilità dello strumento), ma talvolta può accadere il contrario! Si stabilisce allora che l'incertezza nelle misure è data dal maggiore tra questi due errori*"
(prof. Aurelio Agliolo Gallitto, Dipartimento di Fisica, Università di Palermo, http://portale.unipa.it)

Quando poi si è scelto quale tipo di incertezza (si trova spesso scritto, impropriamente: di *errore* ☹) si vuol scrivere accanto alla media delle misure, sarebbe bene **indicare questa scelta ESPLICITAMENTE**!

Vediamo un **ESEMPIO**.
40 misurazioni del periodo T di oscillazione di un pendolo hanno fatto registrare questi valori (in secondi):

4,80	4,82	4,84	4,83	4,79	4,83	4,86	4,86	4,82	4,83	4,87	4,88	4,87	4,89
4,83	4,75	4,86	4,82	4,84	4,87	4,81	4,78	4,85	4,86	4,84	4,79	4,84	4,88
4,85	4,80	4,84	4,85	4,89	4,85	4,83	4,79	4,84	4,81	4,85	4,84		

- La **media** delle misure è stata quindi $4{,}83625$, arrotondata a $4{,}84$
- la **semidispersione** è stata $0{,}07$ per cui potremo scrivere, tenendo conto di essa, $\mathrm{T} = 4{,}84 \pm 0{,}07$
- lo **scarto assoluto medio ("errore medio")** è stato $0{,}0243125$ arrotondabile a $0{,}02$ o a $0{,}024$ per cui, tenendo conto di esso, $\mathrm{T} = 4{,}84 \pm 0{,}02$ o in alternativa $\mathrm{T} = 4{,}836 \pm 0{,}024$
- lo **scarto quadratico medio** è stato $s = 0{,}0309586\ldots$ arrotondato a $0{,}03$ da cui $\mathrm{T} = 4{,}84 \pm 0{,}03$ (verifica che la percentuale dei valori compresi fra $4{,}84 - 0{,}03$ e $4{,}84 + 0{,}03$ non si discosta molto dal 68%!)

Come si vede,
l'intervallo, intorno alla media, che si utilizza per esprimere il valore di una grandezza, dipende dal modo col quale viene espressa l'incertezza; e la corretta interpretazione della scrittura "… ± …" sarà legata alla conoscenza del significato delle varie quantità δ, d, a, s, $s/\sqrt{n}$ … tenendo in debito conto il numero di misure effettuate.

ERRORI RELATIVI / INCERTEZZE RELATIVE
(♥ noiosamente ribadiamolo: è brutta consuetudine ☹ della letteratura scientifica scrivere, tendenzialmente, la parola "errore" anche nei casi in cui il termine corretto sarebbe "incertezza")

L' "errore/incertezza relativo/a" è il quoziente, il rapporto, fra un errore/incertezza (di qualsiasi tipo!) e il valore della grandezza da misurare (valutato tramite la media delle misure rilevate; se tale valore fosse negativo, si intende di ignorarne il segno, cioè di prenderlo in valore assoluto).

Consideriamo nuovamente le 40 misure del periodo di un pendolo elencate a pag. 414.

- La media delle misure è stata $4{,}83625$, arrotondata a $4{,}84$
  La semidispersione è stata $0{,}07$, da cui la possibilità di scrivere $T=4{,}84\pm 0{,}07$.
  Dunque l'incertezza assoluta viene qui valutata in $0{,}07$:
  bene, l'incertezza relativa sarà allora all'incirca di $\frac{0{,}07}{4{,}84}\approx 0{,}014$.
  In forma percentuale, l'incertezza relativa è (circa) dell' $1{,}4\,\%$
- L'errore medio è stato $0{,}0243125$ arrotondato a $0{,}02$ per cui, tenendo conto di esso, $T=4{,}84\pm 0{,}02$:
  l'errore medio relativo è (circa) $\frac{0{,}02}{4{,}84}\approx 0{,}004$, e l'errore medio relativo percentuale circa dello $0{,}4\%$
- Lo scarto quadratico medio è stato $s=0{,}0309586...$ arrotondato a $0{,}03$ da cui $T=4{,}84\pm 0{,}03$
  quindi lo **scarto quadratico medio relativo** - detto, come sappiamo, "**coefficiente di variazione**" - è (circa) $\frac{0{,}03}{4{,}84}\approx 0{,}006$, e lo scarto quadratico medio relativo percentuale è all'incirca dello $0{,}6\%$

L'incertezza relativa può essere impiegata per confrontare la precisione di misure di quantità diverse.

Ad esempio, se nella misura dell'altezza di una parete A c'è l'incertezza di 10 cm
mentre nella misura dell'altezza di un'altra parete B l'incertezza è di 20 cm,
non possiamo affermare che la misura di A sia più precisa di quella di B
se non conosciamo quanto valgono, all'incirca, le altezze di A e di B …
Poniamo che A sia una casa a due piani alta pressappoco 6 metri e B un grattacielo di circa 130 metri:
l'incertezza relativa su A sarà di $\frac{0{,}10}{6}\approx 0{,}017$ mentre l'incertezza relativa su B di $\frac{0{,}20}{130}\approx 0{,}0015$ (meno della decima parte della precedente!), quindi in questo caso va considerata di gran lunga più precisa la misura di B.

GLI ERRORI "SISTEMATICI"

Nel valutare la misura di una grandezza fisica,
oltre agli **errori "CASUALI" (detti anche "ACCIDENTALI" o "STATISTICI")**
(ossia: oltre agli errori **legati a circostanze *imprevedibili* e mai completamente controllabili, le quali possono influire sul risultato della misura *ora per difetto, ora per eccesso***),
si possono commettere anche errori cosiddetti **SISTEMATICI**.

Questi **influiscono sempre per difetto o sempre per eccesso sul valore rilevato**, e derivano:

- ❑ dall'**inadeguatezza dello strumento di misura** (esempi: un orologio che "ritardi", un termometro che con la propria temperatura vada a modificare in modo sensibile la temperatura dell'oggetto in esame …)
- ❑ dall'**uso non appropriato di tale strumento** (es.: dimenticarsi di "azzerarlo", quando ciò sia necessario)
- ❑ da **applicazione di leggi sbagliate o metodi sbagliati di indagine**
  (ad esempio cercare di determinare la profondità di un pozzo lasciandovi cadere una pietra e annotando dopo quanti secondi si sente "splash", per poi utilizzare la formula nota che regola spazi e tempi nella caduta dei gravi … ma senza tener conto che il suono dell'impatto con l'acqua ci mette a sua volta un certo tempo per salire dal fondo del pozzo alle nostre orecchie).

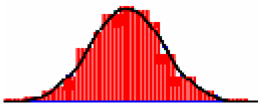
Gli errori sistematici possono essere individuati ed eliminati o perlomeno minimizzati, mentre sugli errori accidentali non possiamo far nulla
(a parte, è ovvio, cercare di effettuare l'operazione di misura con tutta l'attenzione di cui siamo capaci);
l'incertezza legata agli errori accidentali è ineliminabile: può solo essere quantificata coi metodi visti sopra, e ridotta facendo, se possibile, un numero elevato di misure.

Alcuni testi introducono come categoria a sé stante gli "**ERRORI DI SENSIBILITA'** ",
ossia quelli **legati alla sensibilità dello strumento**.
Se misuro la larghezza di un foglio di carta con un righello le cui tacche più ravvicinate siano quelle dei mm,
a ogni misura sarà comunque associata un'incertezza di $0{,}5$ mm (secondo alcuni, di 1 mm)

Gli errori casuali si presentano solo quando sono maggiori della sensibilità dello strumento!!!

ESERCIZI (risposte a pag. 426)

1) Calcola, per il seguente insieme di dati: 0 1 2 2 5
   a) la media ……….. b) la semidispersione ………….. c) lo scarto assoluto medio ……………
   d) la varianza ……….. e) lo scarto quadratico medio (arrotondato a 1 cifra dopo la virgola) ………..

2) VERO O FALSO?
   a) "*Scarto quadratico medio*" e "*deviazione standard*" sono sinonimi
   b) Se effettuo tantissime misure di una grandezza *G*,
   e calcolo la loro media $\overline{x}$ e la loro deviazione standard s,
   nello scrivere $G = \overline{x} \pm s$ io intendo che l'intervallo da $\overline{x} - s$ a $\overline{x} + s$
   ha una probabilità del 68% circa di contenere il vero valore della grandezza
   c) La media $\overline{x}$ fra un numero elevato *n* di misure è una buona approssimazione del valore vero μ
   della grandezza, e se a questo punto faccio *k* misure in più e vado a calcolare la media fra tutte
   le $n + k$ misure, certamente tale nuova media sarà ancora più vicina al vero valore della grandezza
   d) Nella figura  che si riferisce alle ripetute misurazioni di una quantità fisica, le altezze dei rettangoli rappresentano le frequenze
   e) Nella stessa figura di prima, le basi dei rettangoli rappresentano le classi di misura
   f) Lo scarto quadratico medio "corretto" è minore di quello "non corretto"
   g) La funzione "scarto quadratico medio" (non corretto) si indica, nel foglio elettronico, con dev.st.()
   h) Per dimezzare l' "errore standard della media" occorre raddoppiare il numero delle misure

3) Sono state rilevate 625 misure. La media di queste è stata $\overline{x} = 152,4$ e lo scarto quadratico medio $s = 2,5$.
   a) Un intervallo nel quale rientrerà, pressappoco, il 95% delle misure effettuate è quello che va
   da …………… a ……………
   b) Un intervallo di confidenza al 95% per il vero valore della grandezza in esame
   (ossia, un intervallo che ha una probabilità intorno al 95% di contenere il vero valore della grandezza)
   è invece quello compreso fra …………… e ……………

4) La media fra 64 misurazioni di una grandezza risulta essere 173,5 e il loro scarto quadratico medio 2,3.
   a) Determina un intervallo nel quale dovrebbe rientrare pressappoco il 68% di questi 64 dati
   b) Determina un intervallo di confidenza al 68% per il valore della grandezza in esame

5) Misurando 80 volte il tempo di caduta di un grave da una data altezza, in secondi, sono stati trovati i valori:

43,3	41,2	42,5	42,4	43,4	43	44	42	41,6	40,8	42,4	44	43,8	42	43,4	42,2
42,8	41,5	44,5	43,3	42,5	44,2	41,8	42,4	42,2	43,2	42,2	41,8	42,1	43,1	41,7	42,1
42,8	44	43,3	42,7	44,3	44,1	41,4	42,5	42,8	42,8	43,1	42,1	43	42	42,2	42
42,5	41	43,6	43,3	42,9	43,2	42,3	42,9	42,3	41,8	42,2	42,7	41,3	44,4	42,8	42,8
42,1	40,9	43,7	43,6	43,4	43	42,4	44,3	41,7	43,7	42,7	43,4	42,3	42,1	43,3	42

Foglio elet-tro-nico!

   a) Esprimi quel tempo come ... ± ... utilizzando la deviazione standard e arrotondando media e dev. standard a 1 cifra dopo la virgola. Conta il numero di dati tra $\overline{x} - s$ e $\overline{x} + s$ e il numero di quelli tra $\overline{x} - 2s$ e $\overline{x} + 2s$.
   b) Determina un intervallo di confidenza al 95% (cosa significa?) per il valore della grandezza in esame.

6) Clicca sulla freccia per un altro esercizio di questo tipo, con 400 dati già pronti ⇨

7) Si vuole stimare l'età media in cui si presenta una data patologia.
I 400 pazienti seguiti da un famoso centro specializzato hanno contratto la malattia all'età media di 44 anni.
La distribuzione di queste 400 età è *Gauss-like*, con scarto quadratico medio uguale a 10 anni.
Se ne deduce allora che l'intervallo di età che va da … anni a … anni ha una probabilità valutabile intorno al 95% di contenere l'età media di insorgenza della malattia, qualora venisse calcolata sui malati di tutta Italia.
Si può anche dire che, in quel campione di 400 malati, pressappoco il 95% avrà contratto la malattia nell'intervallo di età che va da … anni a … anni;
e siccome il campione, piuttosto numeroso, rispecchia la popolazione generale, tale intervallo di età sarà anche quello entro il quale sviluppa la malattia il 95% circa degli italiani che si ammalano.

8) Dal sito www.regentsprep.org:
Battery lifetime is normally distributed (= segue la distribuzione *normale*, cioè *gaussiana*) for large samples (*sample* = *campione*).
The mean lifetime is 500 days and the standard deviation is 61 days.
What percent of batteries have lifetimes longer than 561 days?

9) Supponi che in una grande città sudamericana sia stata rilevata l'altezza di 420 ragazzi quattordicenni, ottenendo una media di m 1,67 e uno scarto quadratico medio di cm 10.
Allora un intervallo di altezze che ha il 95% di probabilità di contenere l'altezza media di tutti i ragazzi di quell'età, residenti in quella città, è quello che va da ……. a …….

10) I moderni test per attribuire il cosiddetto "quoziente di intelligenza" (Q.I.) sono progettati in modo che nella curva, approssimativamente gaussiana, ottenuta disponendo in orizzontale i vari punteggi realizzabili e in verticale il numero di persone, tutte di una stessa età e, che hanno realizzato quella determinata fascia di punteggio, la media risulti uguale a 100 e la deviazione standard a 15.
In questo modo, pressoché il 95% delle persone della stessa età avrà un Q.I. che si collocherà fra … e …

11) Per una certa popolazione di rane allo stato naturale, si è visto che la lunghezza della vita è distribuita normalmente (cioè, segue una distribuzione gaussiana) con media 10 anni e deviazione standard di 3 anni (www.cli.di.unipi.it).
Quale percentuale di queste rane sopravvive oltre i 16 anni?

12) Sono state effettuate solo 5 misure, che hanno fornito gli esiti seguenti: 85 86,5 85,5 88 86
Se vogliamo esprimere il valore della grandezza con una scrittura del tipo ... ± ... , come faremo?

13) Misurando ripetutamente una grandezza sono stati trovati i valori
2,60 2,59 2,58 2,59 2,59 2,54 2,58.
a) Esprimi quella grandezza come ... ± ... utilizzando la semidispersione.
b) Se scriviamo $G = \overline{x} \pm d$, dove d è la semidispersione,
in generale siamo sicuri che tutte le osservazioni effettuate rientrino fra $\overline{x} - d$ e $\overline{x} + d$?

14) In un sito Internet troviamo che la misura di un dato tempo è $(3{,}27 \pm 0{,}02)\,s$ e che la misura di una data velocità è $(24{,}4 \pm 0{,}3)$ $(mean \pm SD)$. Ma nessuna delle due scritture è scientificamente corretta: perché?

15) Stabilisci quale delle due scritture seguenti esprime una misura di velocità più precisa:
$(3{,}24 \pm 0{,}04)\,m/s$ $(mean;\ SD)$; $(40{,}5 \pm 0{,}5)\,m/s$ $(mean;\ SD)$

16) Esprimendo un tempo come $(8{,}0 \pm 0{,}2)\,s$, qual è l'incertezza relativa percentuale?

17) Se si prendono i due insiemi di dati seguenti: a) 0 3 3 3 6 b) 0 1 2 3 4 per confrontarli, onde stabilire se i dati sono più "sparpagliati" nel primo caso o nel secondo, cosa occorrerebbe calcolare?

18) Lo scarto quadratico medio di n misure è risultato uguale a $2{,}0$ e calcolando l'errore standard della media si è ottenuto $0{,}1$. Quante misure sono state effettuate? …………

19) La tabella qui a destra riporta, in tre casi, il valore di una misura con a fianco l'incertezza da cui questo dato è affetto.
Stabilisci quale delle tre misure può essere considerata la più precisa.

a)	1,25	0,05
b)	10,0	0,3
c)	0,0040	0,0001

20) Una ditta produce camomilla in bustine da 5 grammi.
Si vuole controllare che non troppe bustine abbiano un peso sensibilmente diverso dal valore ottimale.
Pesando 40 bustine prodotte consecutivamente da un macchinario, si trovano i seguenti valori in grammi:

4,89	5,21	5,20	4,76	4,78	5,16	4,84	4,78	4,86	4,88
5,04	5,26	4,74	5,14	4,88	4,82	4,80	5,08	5,20	5,18
5,03	5,18	4,81	4,77	5,20	5,19	5,25	4,75	4,77	4,78
4,90	4,80	4,86	5,18	4,85	4,87	5,05	5,21	5,11	4,82

Quali sono la media e la deviazione standard di questo campione?
Se si è osservato che i dati in esame presentano una *Gauss-like distribution*, delle circa 48000 bustine prodotte in una giornata lavorativa, quante si può presumere che andranno a pesare non più di 4,60 grammi?

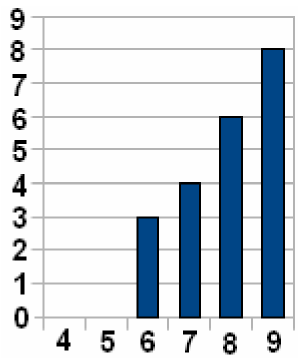
21) SOLO ALCUNE DISTRIBUZIONI SONO GAUSSIANE O *GAUSS-LIKE* , ALTRE NO!!!

Se consideriamo ad esempio la distribuzione dei *pesi delle persone*, o la distribuzione dei *tempi di attesa* dei clienti in una filiale bancaria, capiremo che *le differenze rispetto alla distribuzione normale sono notevoli.*

a) Spiega in che senso queste due distribuzioni, se confrontate con la normale, presentano una "coda verso destra" nella campana.

❑ Nelle **distribuzioni "non normali"**
non si ha necessariamente la coincidenza fra media, moda, mediana.

b) Per fare un esempio semplice, quanto valgono media, mediana e moda nella distribuzione della figura a destra, relativa ai voti dati in una classe di 21 studenti da un insegnante generosissimo? ……. ….... …....



12. ARROTONDAMENTI E CIFRE SIGNIFICATIVE

E' assai comune nella vita quotidiana fare uso di *approssimazioni* di un valore "vero": quando dico "il mio appartamento è di $70\,m^2$", oppure "sono le 11 di sera", o "vado in ferie in un paesino di 200 abitanti", io per l'appunto "approssimo", "arrotondo", e in tutti questi esempi è evidente che lo faccio perché, in simili contesti, non ho bisogno di una precisione più elevata.
Ma anche nelle Scienze sperimentali, ad esempio in Fisica, l'approssimazione di valori numerici è la norma.

LA REGOLA PER ARROTONDARE

La **REGOLA** che si applica per l'arrotondamento di un numero è la seguente.

♪ **Se vengono trasformate in "0" tutte le cifre a partire da una certa cifra e verso destra, quando la prima cifra da trasformare in "0" è 0, 1, 2, 3 o 4, allora nell'arrotondamento la cifra precedente resta invariata;**

♫ **se invece la prima cifra da trasformare in "0" è 5 (ma vedi NOTA), 6, 7, 8 o 9, allora nell'arrotondamento la cifra precedente viene aumentata di un'unità.**

Esempi: l'arrotondamento di 12328 alle centinaia è 12300; quello di 0,1372 ai centesimi è 0,14

NOTA: l'arrotondamento "del banchiere" (banker's rounding, o round-to-even method)

Se la **prima cifra da mutare in 0** è **5**, e tale cifra è **l'ultima del numero**, oppure è **seguita solo da zeri**, allora il passaggio al "valore più vicino" potrebbe essere fatto indifferentemente per difetto o per eccesso, perché ad esempio il numero 1,235 ha la stessa distanza sia da 1,23 che da 1,24;
per questo motivo, nel caso in cui i numeri da sottoporre ad arrotondamento siano tanti, c'è chi preferisce procedere in modo un poco diverso dalla *regola* che abbiamo illustrato, ossia:

se la cifra che precede il 5 è pari, la si lascia invariata, mentre se è dispari, la si aumenta di un'unità.

In tal modo le approssimazioni per difetto e per eccesso così effettuate tenderanno a "bilanciarsi" (sui valori arrotondati secondo questa convenzione, metà circa lo saranno per difetto e metà per eccesso), e l'insieme di dati risentirà il meno possibile, globalmente, delle modifiche apportate.

Per esempio, volendo arrotondare ai centesimi	3,875	3,645	3,735	3,865
si scriverà rispettivamente	3,88	3,64	3,74	3,86

Col "banker's rounding", **l'ultima cifra del numero arrotondato sarà sempre pari**! (*even* = pari)

LE CIFRE SIGNIFICATIVE NELLE SCIENZE SPERIMENTALI

Nelle scienze sperimentali è frequentissimo avere a che fare con numeri dei quali conosciamo con certezza alcune cifre (le prime a sinistra), ma non tutte.

Sono allora "significative" tutte le cifre certe del numero, più la prima cifra incerta.
Questo come idea generale: occhio tuttavia alle specificazioni che seguono.

- Tutte le cifre diverse da 0 sono significative. Ad es., la misura di tempo 11,27 s ha 4 cifre significative.
- Gli 0 iniziali NON sono significativi. Ad esempio, la lunghezza 0,0000245 m ha 3 cifre significative.
- Gli 0 compresi fra cifre non nulle sono significativi.
  4,05 m/s è una velocità espressa con 3 cifre significative.
- Gli 0 finali vanno scritti soltanto se sono significativi, cioè corrispondono alla precisione effettivamente raggiungibile dallo strumento di misura.

  Mi spiego: cm 15,7 può denotare una misura rilevata con uno strumento che ha la precisione dei millimetri, mentre cm 15,70 significherà che lo strumento usato è in grado di apprezzare anche i decimi di millimetro.

  Ancora:
  scrivendo m 1350 per indicare una profondità marina, sottintendo che anche lo 0 finale sia significativo, ossia dichiaro di aver utilizzato una tecnica di misura che mi permetteva di valutare anche il singolo metro.
  Supponiamo invece che già la cifra 5 sia incerta
  (cioè, che le misurazioni effettuate non andassero oltre la precisione dei 10 metri):
  bene, dovrei allora scrivere $1,35 \cdot 10^3$ metri.
- Scrivere il numero in **NOTAZIONE ESPONENZIALE** permette di vedere bene le cifre significative (sono tutte e sole quelle del moltiplicatore della potenza di 10).

  Esempi:
  $0,000107 = 1,07 \cdot 10^{-4}$ (3 *cifre significative*) $\quad$ $5,4 \cdot 10^6$ (2 *cifre significative*)
  $5,40 \cdot 10^6$ (3 *cifre sign.*; *scrivendo così, si evidenzia che anche lo* 0 *è significativo* : 4 *è certa*, 0 *è incerto*)

QUALORA SI SIA FATTO UN CERTO NUMERO DI MISURE PER UNA DATA QUANTITÀ, IL VALORE DELLA QUANTITÀ IN ESAME SI ESPRIMERÀ FACENDO LA MEDIA $\overline{x}$ DELLE MISURE TROVATE, POI SCRIVENDO CHE LA GRANDEZZA IN GIOCO VALE $\overline{x} \pm \Delta x$, dove quel Δx è l'INCERTEZZA (di solito si trova scritto impropriamente: l'ERRORE) che associamo al valore $\overline{x}$, incertezza data dalla semidispersione, oppure dallo scarto quadratico medio o da un suo multiplo, ecc., come abbiamo spiegato nel paragrafo precedente.

Il simbolo Δ ("delta") è sovente utilizzato, in matematica, per indicare "differenza".
Ad es., fra due persone che hanno risp. 15 anni e 47 anni, c'è una differenza di età "delta *e*": $\Delta e = 47 - 15 = 32$.
Se considero, in Fisica, due istanti di tempo successivi t_1 e t_2, nei quali la velocità di un corpo è risp. v_1 e v_2, allora nell'intervallo di tempo $\Delta t = t_2 - t_1$ l'incremento di velocità ($>, <$ o $= 0$) è dato da $\Delta v = v_2 - v_1$.

D'altra parte, **sia la media che l'incertezza subiscono sempre un ARROTONDAMENTO. Vediamo come**.
Andiamo a riprendere i dati sul periodo del pendolo. Le 40 rilevazioni avevano fornito i valori (in secondi):

4,80	4,82	4,84	4,83	4,79	4,83	4,86	4,86	4,82	4,83	4,87	4,88	4,87	4,89
4,83	4,75	4,86	4,82	4,84	4,87	4,81	4,78	4,85	4,86	4,84	4,79	4,84	4,88
4,85	4,80	4,84	4,85	4,89	4,85	4,83	4,79	4,84	4,81	4,85	4,84		

E' evidente che si era utilizzato un dispositivo in grado di apprezzare i *centesimi di secondo*.
La media di queste misure è 4,83625, e lo scarto quadratico medio 0,0309586... . Bene!

Nelle scienze sperimentali di solito si osserva la prassi seguente:

a) L'INCERTEZZA Δx VIENE SEMPRE ARROTONDATA IN MODO CHE CONSERVI UNA CIFRA SIGNIFICATIVA SOLTANTO O AL MASSIMO DUE CIFRE SIGNIFICATIVE SE LA PRIMA DI ESSE È 1 (NOTA)

b) DOPODICHE' LA MEDIA DELLE MISURE SI ARROTONDA IN MODO CHE LA SUA CIFRA PIÙ A DESTRA (= LA CIFRA MENO SIGNIFICATIVA) ABBIA LO STESSO POSTO DECIMALE DELLA CIFRA MENO SIGNIFICATIVA PRESENTE NELL'INCERTEZZA Δx

Insomma, vanno bene $25,7 \pm 0,3$; 178 ± 4; $8,54 \pm 0,07$;
$0,483 \pm 0,016$ (*notare qui la cifra in più nell'incertezza*!)
ma *non* andrebbe bene invece $4,197 \pm 0,05$ oppure $27 \pm 0,4$

Di conseguenza, nel caso del pendolo da noi considerato,
a) **arrotonderemo l'incertezza**: 0,0309586... $\to$ 0,03 **in modo che conservi una sola cifra non nulla**;
b) **poi arrotonderemo la media** 4,83625 $\to$ 4,84 **in modo che la sua cifra più a destra abbia la stessa posizione decimale della cifra più a destra dell'incertezza** (nel nostro caso, i centesimi).
E scriveremo in definitiva che il periodo del nostro pendolo è di secondi $T = 4,84 \pm 0,03$.
Ribadiamolo: 0,03 è qui lo scarto quadratico medio, e il suo significato è di affermare che circa il 68% delle misure effettuate si trova nell'intervallo che ha centro la media e raggio 0,03 e che … ecc. ecc.

- Trovo come media 42,625 e come scarto quadratico medio 0,418?
  Bene, allora arrotondo lo scarto quadratico medio a 0,4 (in modo che rimanga 1 sola cifra significativa) e a questo punto arrotondo pure la media a 42,6 scrivendo il valore della grandezza come $42,6 \pm 0,4$
- Trovo come media 528,25 e come scarto quadratico medio 2,781?
  Bene, allora arrotondo lo scarto quadratico medio a 3 (in modo che rimanga 1 sola cifra significativa) e a questo punto arrotondo pure la media a 528 scrivendo il valore della grandezza come 528 ± 3
- Trovo come media 2,208 e come scarto quadratico medio 0,0331?
  Arrotondo allora lo scarto quadratico medio a 0,03 (in modo che rimanga 1 sola cifra significativa) e a questo punto arrotondo pure la media a 2,21 scrivendo il valore della grandezza come $2,21 \pm 0,03$
- Trovo come media 1,5257 e come scarto quadratico medio 0,0143?
  Arrotondo lo sc. q. m. a 0,014 (ho deciso di tenere 2 cifre significative perché la prima di esse è 1) e a questo punto arrotondo pure la media a 1,526 scrivendo il valore della grandezza come $1,526 \pm 0,014$
- Ho fatto poche misure. La loro media è 10584 e la loro semidispersione è $30 = 3 \cdot 10$.
  La semidispersione ha già una cifra significativa soltanto: va bene così com'è. Ma allora devo arrotondare la media alle decine, e scrivere il valore come 10580 ± 30 o meglio come $(1058 \pm 3) \cdot 10$

NOTA - *Non tutti sono concordi. Noi faremo così, ma alcuni accettano nell'incertezza fino a 2 cifre significative. Altri suggeriscono di usare due cifre significative se la prima cifra è bassa (c'è chi dice 1 o 2, c'è chi dice 1, 2, 3 o 4), altrimenti una. In effetti, se la prima cifra è piccola, eliminare con l'arrotondamento la seconda porterebbe ad una perdita di precisione ritenuta eccessiva anche per un'incertezza. Ma occorre trovare sempre un buon compromesso fra una ragionevole precisione, da una parte, e l'immediata leggibilità della scrittura, dall'altra.*

Supponiamo che un dato sperimentale *non* venga presentato sotto la forma $x \pm \Delta x$, ossia che **non venga specificata nessuna incertezza**: allora si intende che l'**incertezza** sia **implicita nell'ultima cifra**.
Il guaio è che tutto ciò non viene interpretato universalmente allo stesso modo!
Ad esempio, per alcuni $12,3$ è da leggersi come $12,3 \pm 0,05$ ossia $12,25 < x < 12,35$;
per altri, $12,3$ va letto come $12,3 \pm 0,1$ ossia $12,2 < x < 12,4$ ☹

ALTRI ESEMPI

$2,47 \pm 0,03241$ Qui l'incertezza non va bene, va riscritta con una sola cifra significativa: $2,47 \pm 0,03$

$48,57 \pm 0,3$ Qui è il valore che va riscritto. La scrittura dev'essere corretta in $48,6 \pm 0,3$ in maniera che l'ultima cifra della grandezza e l'ultima cifra dell'incertezza abbiano lo stesso posto decimale.

$3831,7 \pm 20$ Non va. L'incertezza è alle decine, quindi il valore va a sua volta arrotondato alle decine: 3830 ± 20 o meglio $(383 \pm 2) \cdot 10$

$18,79 \pm 0,33$ L'incertezza non va, dobbiamo ridurla a una sola cifra significativa.
Scriveremo $18,8 \pm 0,3$ arrotondando anche il valore della grandezza in modo che la sua ultima cifra a destra abbia ugual posto dell'analoga per l'incertezza.

$18,79 \pm 0,13$ Qui possiamo lasciare l'incertezza così com'è, con 2 cifre significative (quindi la scrittura va bene):
"l'incertezza Δx viene sempre arrotondata in modo che conservi 1 cifra significativa soltanto, *o al massimo due cifre sign. se la prima di esse è 1* (c'è chi dice: se la prima di esse è 'piccola')".
Questa eccezione viene accettata perché, se non si facesse così, in casi simili l'arrotondamento dell'incertezza sarebbe troppo "pesante" se rapportato con l'incertezza stessa.

ESERCIZIO (risposte a pag. 427)

1) Prendi in esame ciascuna delle seguenti scritture, per stabilire se è corretta o no.
In quest'ultimo caso, apporta le modifiche appropriate.

a) $x = 27,88 \pm 0,4$ b) $x = 35,73 \pm 0,42$ c) $x = 2,3 \pm 0,0531$ d) $x = 3,25 \pm 0,14$
e) $x = 7,342 \pm 0,079$ f) $x = 4532 \pm 50$ g) $x = 91,3 \pm 2$ h) $x = 0,50 \pm 0,01$

QUANTE CIFRE LASCIARE NEL RISULTATO DI UN CALCOLO SU DATI INCERTI?

Quando si fa una ADDIZIONE o una SOTTRAZIONE fra numeri che derivano da misurazioni affette da incertezza, il risultato dovrà contenere lo stesso numero di CIFRE DOPO LA VIRGOLA dell'addendo che ne contiene di meno.

Facciamo qualche esempio.

- ❑ $9,57 + 12,3 + 2,001 = 23,871$ ma siccome l'addendo con meno cifre dopo la virgola ne contiene 1 sola, la somma $23,871$ dev'essere arrotondata a $23,9$
- ❑ $9,57 + 12 + 2,001 = 23,571$ però qui l'addendo con meno cifre dopo la virgola non ne ha nessuna per cui dobbiamo arrotondare la somma ottenuta alle unità e scriverla come 24
- ❑ $3,2 + 8,77 = 11,97$ che però dev'essere arrotondato a $12,0$
(e il ",0" va conservato perché comunque la cifra 0 dopo la virgola è significativa)

Quando si fa una MOLTIPLICAZIONE o una DIVISIONE fra numeri che derivano da misurazioni affette da incertezza, il risultato dovrà contenere lo stesso numero di CIFRE SIGNIFICATIVE del termine che ne contiene di meno.

Esempi.

- ❑ $7,081 \cdot 4,32 = 30,58992$
ma per conservare solo 3 cifre significative (quante ne ha il 2° fattore), siamo costretti ad arrotondare a $30,6$
- ❑ $7,4 \cdot 1,43 = 10,582$
e tuttavia dovremo arrotondare in modo che le cifre significative siano solo 2 … scrivendo perciò 11
- ❑ $58,4 : 0,023 = 2539,13...$
… però il risultato non potrà essere scritto con più di 2 cifre significative (quante ne ha il divisore) quindi andrebbe arrotondato a 2500, che tuttavia, scritto così, di cifre significative pare averne quattro …
… risolviamo l'inghippo scrivendo il quoziente in notazione esponenziale, come $2,5 \cdot 10^3$
- ❑ $4,02 \cdot 0,49754 = 2,0001108$ ma qui occorre fare in modo che nel risultato le cifre significative siano soltanto tre, come nel primo fattore. Bene: il risultato andrà allora scritto come $2,00$

ESERCIZIO (risposte a pag. 427)

2) Considera le coppie *x, y* di dati seguenti. L'ultima cifra a destra è incerta.

a)	b)	c)	d)
$x=9{,}5$	$x=32{,}6$	$x=54{,}3$	$x=2{,}355$
$y=2{,}37$	$y=43$	$y=0{,}24$	$y=3{,}2$

E' richiesto di scrivere col numero corretto di cifre I) $x+y$ II) $x-y$ III) xy IV) $\frac{x}{y}$

LA "PROPAGAZIONE" DEGLI ERRORI, O MEGLIO: DELLE "INCERTEZZE"

NOTA

SULLA DISTINZIONE FRA "ERRORE" E "INCERTEZZA"

Cos'è l' ERRORE? E' la differenza (presa in valore assoluto) fra il valore approssimato, o il valore ricavato da una misura, e il valore vero.

Occhio ... SI È PURTROPPO AFFERMATA L'INFELICE CONSUETUDINE DI CHIAMARE SBRIGATIVAMENTE E IMPROPRIAMENTE "ERRORE" ANCHE CIÒ CHE IN REALTÀ DOVREBBE ESSERE DENOMINATO "INCERTEZZA". ☹

Il termine "INCERTEZZA" denota

- **in senso stretto,**
  una "maggiorazione dell'errore":
  se il valore approssimato o rilevato è x, insomma,
  e l'incertezza, intesa in questo modo, è k,
  allora il vero valore sarà compreso fra $x-k$ e $x+k$;
- **in un'accezione più generale, il grado di indeterminazione cui è soggetto il valore che viene attribuito a una data quantità.**

Siano a, b due grandezze, e sia G una terza grandezza che derivi da un'operazione aritmetica su a, b. Allora:

- **L' incertezza della SOMMA $G=a+b$**
  è la somma delle incertezze da cui sono affetti gli addendi: $\boxed{\Delta G=\Delta a+\Delta b \ \text{ se } \ G=a+b}$

  Di solito questa regola viene enunciata impropriamente così ☹ :
  L'errore della somma è uguale alla somma degli errori degli addendi
- **La stessa identica regola vale per la differenza: l' incertezza della DIFFERENZA $G=a-b$**
  è la somma delle incertezze da cui sono affetti i termini: $\boxed{\Delta G=\Delta a+\Delta b \ \text{ se } \ G=a-b}$
- **L'incertezza del PRODOTTO $G=ca$ DI UN NUMERO COSTANTE $c>0$ PER UNA GRANDEZZA**
  è il prodotto del numero fisso per l'incertezza della grandezza: $\boxed{\Delta G=c\Delta a \ \text{ se } \ G=ca}$
- **L' *incertezza relativa* (OCCHIO! RELATIVA, questa volta, non assoluta!) del PRODOTTO $G=a\cdot b$**
  è la somma delle *incertezze relative* dei fattori: $\boxed{\frac{\Delta G}{G}=\frac{\Delta a}{a}+\frac{\Delta b}{b} \ \text{ se } \ G=a\cdot b}$

  Di solito questa regola viene enunciata impropriamente così ☹ :
  L'errore *relativo* del *prodotto* è la *somma* degli errori *relativi* dei fattori
- **Del tutto analoga a quella sul prodotto, e come essa basata sulle incertezze *relative*,**
  è la regola per il QUOZIENTE $G=\frac{a}{b}$: $\boxed{\frac{\Delta G}{G}=\frac{\Delta a}{a}+\frac{\Delta b}{b} \ \text{ se } \ G=\frac{a}{b}}$
- **Per la POTENZA $G=a^n$:** $\boxed{\frac{\Delta G}{G}=n\frac{\Delta a}{a} \ \text{ se } \ G=a^n}$ valida anche se n è frazionario, ossia con le radici! $\sqrt[3]{x^2}=x^{\frac{2}{3}}$ $\sqrt{x}=x^{\frac{1}{2}}$

ESERCIZIO (risposte a pag. 427)

3) Considera le coppie *x, y* di dati seguenti; per ciascun dato è specificata l'incertezza da cui è affetto.

a)	b)	c)	d)
$x=9{,}5\pm0{,}5$	$x=32{,}6\pm0{,}4$	$x=54{,}3\pm0{,}8$	$x=2{,}355\pm0{,}006$
$y=2{,}37\pm0{,}03$	$y=43\pm1$	$y=0{,}24\pm0{,}02$	$y=3{,}2\pm0{,}1$

Determina le incertezze *assoluta* e *relativa* di: I) $x+y$ II) $x-y$ III) xy IV) x/y V) x^4 VI) $\sqrt{x}$

NOTA - In casi come questi, si fanno i calcoli intermedi con la totalità delle cifre; *soltanto alla fine* si arrotonda.

13. RISPOSTE AGLI ESERCIZI

RISPOSTE agli esercizi delle pagg. 360-361 (CONCETTI INTRODUTTIVI)

2) a) quant. discr. b) qual. sconn. c) quant. discr. d) quant. cont. e) quant. discr. f) qual. ord. g) qual. sconn.
h) quant. discr., anche se poi è opportuno che i dati vengano ripartiti in "classi" (es.: *meno di 5000* abitanti …)

3e) "Mai": 33; 0,66; 66%; …

6)

Numero figli	Freq. ass.	Freq. rel.	Freq. perc.
0	8	8/40 = 0,2	20%
…	…	…	…

7) Molto+Abbastanza+Poco+Pochissimo = $12+5+5+2=24$; $24 = 80/100\ x$ (x = n° totale) da cui $x = 30$
e perciò Moltissimo = 30 – 24 = 6; Freq. rel. (Moltissimo) = 0,2; Freq. rel. (Molto) = 12/30 = 0,4; ecc.

8) La somma delle freq. rel. è sempre 1.
La freq. rel. della modalità rimanente è perciò $1-0{,}35-0{,}4-0{,}2=0{,}05$ che corrisponde a una perc. del 5%.

9) a) F b) F c) F d) F e) F

11) Ad esempio, per classi di 1 voto:

Classe di freq.	Freq. ass.
$4 \le x < 5$	1
…	…

… e per classi di ½ voto:

Classe di freq.	Freq. ass.
$4 \le x < 4{,}5$	0
…	…

12) Ad esempio, per classi di 7 giorni:

Classe di freq.	Freq. ass.	Freq. rel.	Freq. perc.
Da 1 a 7 gg.	8	0,17	17%
…	…	…	…

13) a) Sì b) Sì c) Sì d) No e) No (anche se si potrebbe inserire un rarissimo "4 o più") f) Sì g) No h) Sì

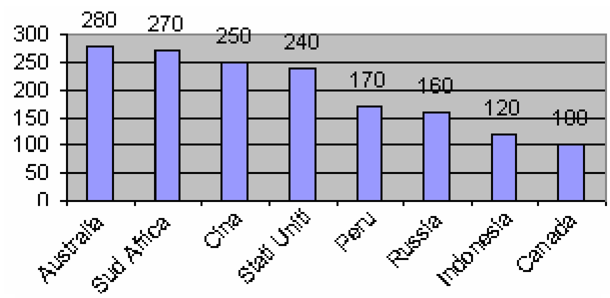
RISPOSTE agli esercizi delle pagg. da 380 a 385 (RAPPRESENTAZIONI GRAFICHE)

1) Ad esempio, arrotondando all'intero, $\alpha°(Africa) = 996/6776 \cdot 360° \approx 53°$

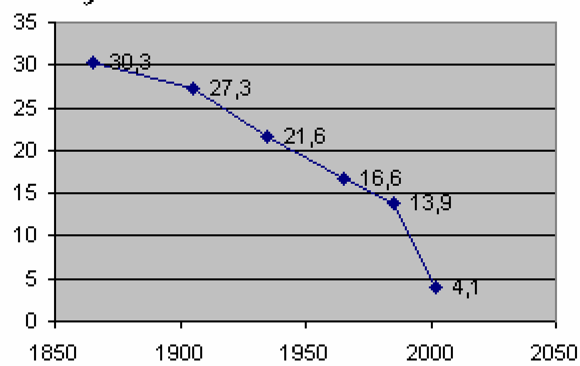
2) Ad esempio, arrotondando all'intero, $\alpha°(Ossigeno) = 46{,}6/100 \cdot 360° \approx 168°$

3) Si potrebbe prevedere una "fetta" unica per tutti gli elementi presenti in percentuale < 1%, o in tracce.

4)



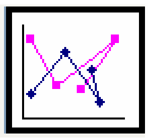
6)



← Serie storica tracciata con Excel 2003

a) scegliendo Dispers. (XY)

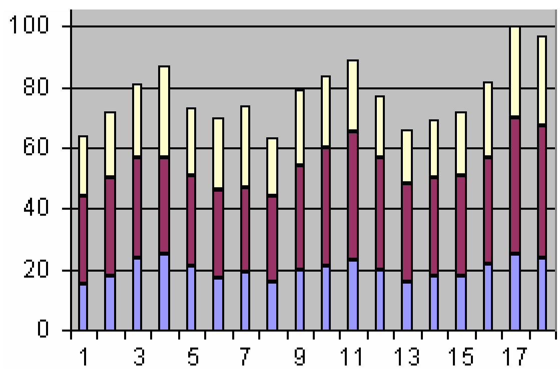
b) poi "Dispersione con coordinate unite da linee"



c) cliccando su "Etichette dati", quindi sul quadratino accanto a "Valori (Y)"

7) Si potrebbe, ad esempio, pensare alle classi:
da 45 km/h compresi a 50 km/h esclusi;
da 50 compresi a 55 esclusi; ecc.
Per contare il numero di dati di ciascuna classe, puoi ricorrere ad un uso accorto della funzione CONTA.SE, come spiegato a pag. 377.

8) In questo caso, la rappresentazione più "espressiva" è senz'altro quella del tipo "Istogramma in pila" →



9) a) 49 b) 7; 0,14 circa; 14% circa c) circa il 61% 10) 5; 0,19 (con arrotondamento); 19% (circa)

12) b) 10 080 000 circa c) perché diminuisce la natalità (fortissima discesa in 3ª colonna) ma simultaneamente la popolazione è in aumento (principalmente in quanto si vive mediamente più a lungo)

20) Guardando solo l'ideogramma, $\frac{2,5}{3}\cdot 100 \approx 83\%$. Coi numeri, più precisamente, $\frac{1212000}{1550000}\cdot 100 \approx 78\%$

21) Onestà e competenza = 42% circa … 22) 18 eccellenti (media non inferiore a 9) …

23) Ad esempio, il rettangolo più a sinistra ha base 4 e altezza 149,5

RISPOSTE agli esercizi delle pagg. da 400 a 405 (INDICI DI POSIZIONE)

1) Media leggerissimamente superiore a 6,7 (6,7045…); due "mode": 6,5 e 7,5; mediana = 6,75
2) Media leggerissimamente superiore a 1,9 ore (1,903…); mediana = 2; moda = 1,5
3) $\frac{1,78\cdot 5+x}{6}=1,80; \quad 8,90+x=10,80; \quad x=1,90$
4) In generale, no: non coinciderà. Potrebbe eccezionalmente coincidere in casi particolari:
ad esempio, se i gruppi hanno ugual numero di elementi, coincide.
Dimostra questo fatto per un caso particolare: ad esempio, considerando 6 dati $x_1, x_2, x_3, x_4, x_5, x_6$.
Se questo insieme di 6 elementi viene spezzato in 2 gruppi di 3 elementi ciascuno,
oppure in 3 gruppi di 2 elementi ciascuno, la media generale sarà senz'altro uguale alla media delle medie.
5) Non è possibile rispondere basandosi solo su questi dati!
Bisognerebbe infatti sapere quanti sono gli abitanti, o più precisamente gli aventi diritto al voto,
in ciascuna delle due regioni, o almeno qual è il rapporto fra il loro numero nella regione A e nella regione B.
6) Stessa identica media! (25' 28")
7) a) $Media \approx 14,24$ giorni
b) La media *per classi* di 7 gg. non differisce di molto dalla media "normale": si trova infatti $\approx 14,04$ gg.
8) a) spesa minima 7,05 euro, spesa massima 254,50 euro b) spesa media individuale $\approx 50,53$ euro
e) La media "esatta" e quella "per classi" di 10 euro differiscono di pochissimo:
facendo la media per classi si trova infatti esattamente 50,50 euro
9) Beh, la media esatta, no, ma la media per "classi", prendendo per ciascuna classe il suo valore centrale, sì.
E, come abbiamo precedentemente visto su un paio di esempi e come si potrebbe verificare,
la media così calcolata è una buona approssimazione della vera media.
Si ottiene nel nostro caso una media prossima a 9,78 euro.
10) Poco più di 13 euro 11) Media $\approx 8,59$ euro (arrotondando ai centesimi); mediana = 7,5; moda = 5
12) Media leggermente sup. a 11,4 anni; mediana = 11; moda = 12 13) Media $\approx 4,2$; mediana = 4; moda = 4
14) Voto finale da 0 a 10: ad esempio, $Voto\ Paolo=\frac{8\cdot 1+7\cdot 1,5+7\cdot 0,8+7\cdot 2+8\cdot 0,5}{1+1,5+0,8+2+0,5}\approx 7,26$

Per Serena e Martina si ottengono, arrotondando sempre a due cifre dopo la virgola, le medie seguenti:
$Serena \approx 7,91$; $Martina \approx 7,62$. Certo, l'insegnante dovrà poi procedere a un arrotondamento ulteriore …

Voto finale dal 2 al 10: si tratta
I) di restringere la fascia da 0 a 10, in modo che al suo posto si abbia una fascia da 0 a 8
II) poi di traslare verso l'alto di 2 unità: $v'=\frac{4}{5}v+2$, dove si determina prima v col metodo precedente.

15) $media_{min} \approx 4,14$
16) Sì, perché conoscendo la media e il numero dei dati è possibile risalire alla somma dei dati.
La somma di tutti i punteggi della II A è $7,25\cdot 22=159,5$ e quella dei punteggi di II B è $7,8\cdot 28=218,4$.
Quindi la somma dei punteggi riunendo insieme i 2 gruppi è $377,9$ e la media gen. è $377,9/50=7,558$
17) Cambierebbe la media aritmetica, ma mediana e classe modale resterebbero inalterate 18) 7,44
22) Sommando i tempi si ottiene 11h 29' 20", e sommando le distanze 396,5 km. La velocità media
sull'intero tragitto è perciò $\frac{396,5 \text{ km}}{11h\ 29'\ 20''}\approx\frac{396,5 \text{ km}}{11,49 \text{ h}}\approx 34,51$ km/h [11h 29' 20" = 41360" = = 41360:3600 h ≈ 11,49 h]
23) $Media \approx 25,47$; $mediana = 25$; $moda = 25$
24) $Media = 57$ km/h; $mediana = 55,5$; $moda = 55$. Per classi (da 45 km/h a 49, da 50 a 54, …):
$2\ classi\ modali,\ 50 \le v \le 54\ e\ 55 \le v \le 59$; $classe\ mediana\ 55 \le v \le 59$; $media\ per\ classi = 57$ km/h
25) Vedrai che uscirà come media un valore molto prossimo a 3,5.

26) $Media = 3,8;\ \ mediana = moda = 4$ 27) $Media\ (per\ classi) \approx 2,60;\ \ cl.\ mediana = cl.\ modale:\ 2,5 \le x < 3$

29) $\approx 3,9$

30) Il problema sta nella genericità di quel "6 O PIÙ". Supponendo che il "6 o più" sia un 6, si ottiene nel 1961 una media di 3,48 componenti per famiglia; questa è dunque una stima per difetto. Per il numero approssimativo totale dei residenti basta moltiplicare il numero medio di componenti per famiglia per il numero delle famiglie, disponibile sull'ultima riga (che dà il numero di *migliaia* di famiglie relativo a quell'anno). Per il 2001, data la bassa percentuale di famiglie con "6 o più" componenti, la media calcolata sostituendo quel "6 o più" con 6 è più attendibile rispetto all'analoga per il 1961. Si ottiene, per il 2001, media $\approx 2,59$ componenti per famiglia e numero totale residenti vicino a 56.500.000.

31) Qui scriveremo i risultati arrotondandoli a 2 cifre decimali (se ne avevano più di 2). La REGOLA che applicheremo per l'arrotondamento di un numero è la seguente.

♪ Se vengono trasformate in "0" tutte le cifre a partire da una certa cifra e verso destra, quando la prima cifra da trasformare in "0" è 0, 1, 2, 3 o 4, allora nell'arrotondamento la cifra precedente resta invariata; es. $8{,}137105 \to 8{,}1$

♫ se invece la prima cifra da trasformare in "0" è 5, 6, 7, 8 o 9, allora nell'arrotondamento la cifra precedente viene aumentata di un'unità; es. $8{,}16 \to 8{,}2$

a) $M = 4;\ M_G \approx 3{,}35;\ M_A \approx 2{,}60;\ M_Q \approx 4{,}47$; b) $M = 1{,}2;\ M_G \approx 1{,}15;\ M_A \approx 1{,}11;\ M_Q \approx 1{,}26$;

c) $M \approx 0{,}58;\ M_G = 0{,}5;\ M_A \approx 0{,}43;\ M_Q \approx 0{,}66$

32) $x^3 = 30 \cdot 40 \cdot 50\ \ da\ cui\ \ x = \sqrt[3]{30 \cdot 40 \cdot 50} = \sqrt[3]{60000} \approx 39{,}15$ (media geometrica delle dimensioni).

33) La risposta esatta è 150 km/h, ossia la media armonica $v = \dfrac{1}{\dfrac{\frac{1}{100} + \frac{1}{300}}{2}}$.

E' presumibile che il VIP non sia caduto nel tranello di utilizzare la media aritmetica perché conosceva già la risposta a questo quesito o comunque a quesiti simili; o anche perché, con la sua intelligenza "pratica", aveva capito immediatamente che la domanda era stata posta per metterlo in difficoltà, e quindi la risposta più "banale" (media aritmetica, 200 km/h) non poteva essere quella giusta. In ogni caso è stato bravo, e probabilmente non ha sfruttato direttamente la formula per la media armonica, ma ha ragionato in questo modo, dando allo spazio totale un valore "comodo per i calcoli": supponiamo che il percorso complessivo sia di 600 km; per fare i primi 300 ci si mette 3 ore, per fare gli altri 300 ci si impiega 1 ora. 4 ore in totale, 600 km, da cui 600:4 = 150 kilometri all'ora.

34) AH rappresenta la media geometrica di BH e HC, AM la media aritmetica.

35) Perché nel triangolo rettangolo AHM il cateto AH è sempre < dell'ipotenusa AM. Si avrebbe l'uguaglianza $\mathrm{AH} = \mathrm{AM}$ se i due segmenti fossero fra loro sovrapposti, il che avviene quando ABC è isoscele.

36) La media quadratica dei cateti. Infatti, per qualunque coppia di cateti, è (Teorema di Pitagora)
$\sqrt{\mathrm{PA}^2 + \mathrm{PB}^2} = \sqrt{\mathrm{P'A}^2 + \mathrm{P'B}^2} = \sqrt{\mathrm{P''A}^2 + \mathrm{P''B}^2} = \mathrm{AB} = costante$.

Quindi è costante, per ogni coppia di cateti a, b, anche la quantità $\dfrac{\sqrt{a^2 + b^2}}{\sqrt{2}} = \sqrt{\dfrac{a^2 + b^2}{2}}$ (media quadr.)

37) $v = \dfrac{1}{\dfrac{\frac{1}{30} + \frac{1}{35} + \frac{1}{45}}{3}} = \dfrac{3}{\dfrac{35 \cdot 45 + 30 \cdot 45 + 30 \cdot 35}{30 \cdot 35 \cdot 45}} = \dfrac{3 \cdot 30 \cdot 35 \cdot 45}{1575 + 1350 + 1050} = \dfrac{141750}{3975} \approx 35{,}66$ (km/h)

E' la media armonica delle 3 velocità. Osserviamo che la risposta non dipende dalla lunghezza del percorso: se il circuito fosse stato di 5 km, o di 700 metri, avremmo ottenuto il medesimo risultato.

38) $\dfrac{30 + 35 + 45}{3}$ km/h $\approx 36{,}67$ km/h (media aritmetica delle tre velocità)

Osserviamo che la risposta non dipende dal tempo, nel senso che sostituendo a "5 minuti" un altro intervallo di tempo qualsiasi, la velocità media rimarrebbe sempre la stessa.

39) a) 6 km/h b) 1,8 km/h *Rifletti sul motivo della grande differenza rispetto alla risposta a) ...* 40) $\dfrac{24 + v}{2} = 27;\ 24 + v = 54;\ v = 30$ km/h

41) $\dfrac{24 + v + v + v}{4} = 27;\ 24 + 3v = 108;\ 3v = 84;\ v = 28$ km/h *OPPURE*: $\dfrac{s}{t} = \dfrac{\frac{1}{4} \cdot 24 + \frac{3}{4} \cdot v}{1} = 27;\ \frac{3}{4}v = 21;\ v = 28$ km/h

42) $\dfrac{1}{\dfrac{\frac{1}{24}+\frac{1}{v}}{2}}=27;\quad \dfrac{2}{\frac{1}{24}+\frac{1}{v}}=27;\quad \dfrac{2}{\frac{v+24}{24v}}=27;\quad \dfrac{48v}{v+24}=27;\quad 48v=27v+24\cdot 27;\quad 21v=648;\quad v\approx 30{,}86$

43) Nel 1° tratto di 6 km, velocità $=\dfrac{6\text{ km}}{\frac{1}{4}\text{ h}}=24\text{ km/h}\ \rightarrow\ \dfrac{1}{\dfrac{\frac{1}{24}+\frac{1}{v}+\frac{1}{v}+\frac{1}{v}+\frac{1}{v}}{5}}=27;\ \ldots\ v\approx 27{,}87$ $\quad$ *OPPURE*: $\dfrac{s}{t}=\dfrac{6+24}{\frac{1}{4}+\frac{24}{v}}=27$ ecc.

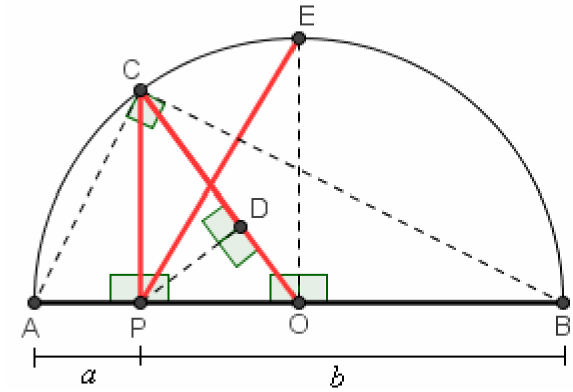
44) Al termine del terzo anno il posseggo il 99,498% di ciò che possedevo inizialmente:
ci ho quindi perso un pochino (leggermente più dello 0,5%)

46) Il tasso di interesse medio annuo è del 22,4745% circa (approssimazione per leggerissimo eccesso).
In sé questo 22,4745 (approssimato) non rappresenta una media di alcun tipo, ma si può dire che 122,4745 (ammontare del debito dopo 1 anno, se la cifra iniziale era 100) rappresenta la media geometrica fra 100 e 150

47) $\cancel{p}\left(1-\dfrac{x}{100}\right)^2=0{,}5\cancel{p};\quad 1-\dfrac{x}{100}=\sqrt{0{,}5};\quad -\dfrac{x}{100}=\sqrt{0{,}5}-1;\quad \dfrac{x}{100}=1-\sqrt{0{,}5};\quad x=100\cdot\left(1-\sqrt{0{,}5}\right)\approx$
$\approx 100\cdot(1-0{,}707)=100\cdot 0{,}293=29{,}3$ $\quad$ La perdita di valore media annua è stata circa del 29,3%

48) $\dfrac{k}{\dfrac{\frac{k}{24}+\frac{k}{40}}{2}}=\dfrac{1}{\dfrac{\frac{1}{24}+\frac{1}{40}}{2}}=\dfrac{2}{\frac{1}{24}+\frac{1}{40}}=\dfrac{2}{\frac{40+24}{24\cdot 40}}=\dfrac{2\cdot 24\cdot 40}{64}=30\ \ (\textit{media armonica})$

49) Traccia innanzitutto CA, CB; $\widehat{ACB}$ sarà di 90°
perché inscritto in una semicirconferenza;
per Euclide II°, o coi triangoli simili, si ha allora
$\mathrm{PC}=\sqrt{\mathrm{AP}\cdot\mathrm{PB}}=\sqrt{ab}=M_G(a,b)$.
Poi: $\mathrm{OC}=r=\mathrm{AB}/2=(a+b)/2=M(a,b)$;
$\mathrm{DC}:\mathrm{PC}=\mathrm{PC}:\mathrm{OC}$ (*PDC simile con OPC*) $\rightarrow$
$\rightarrow \mathrm{DC}=\dfrac{\mathrm{PC}^2}{\mathrm{OC}}=\dfrac{ab}{\frac{a+b}{2}}=\dfrac{1}{\frac{\frac{a+b}{ab}}{2}}=\dfrac{1}{\frac{\frac{1}{a}+\frac{1}{b}}{2}}=M_A(a,b);$
…



RISPOSTE agli esercizi delle pagg. 406-407 (INDICI DI DISPERSIONE)

1) a) I) campo di variabilità = 6 II) scarto ass. medio = 1,6 III) deviaz. st. = 2 IV) coeff. di variaz. = 0,5
b) I) campo di var. = 1 II) scarto ass. medio = 0,32 III) deviaz. st. = 0,4 IV) coeff. di variaz. = 1/3
c) I) c. var. = 3/4 II) scarto ass. medio = 5/18 III) dev. st. = $\sqrt{7/72}\approx 0{,}3118$ IV) coeff. var. $\approx$ 0,5345

2) Senza dubbio è preferibile "scarto assoluto medio" (= la media degli scarti, presi in valore assoluto).
"Scarto medio", per la smania di abbreviare evitando un aggettivo, in realtà pretende che il lettore questo aggettivo lo tenga presente molto bene, perché la media degli scarti "e-basta" sarebbe 0!
"Scarto medio assoluto", se presa alla lettera, vorrebbe dire che
calcolo la media degli scarti (ottenendo 0), poi di questa media faccio il valore assoluto: risultato finale 0.

3) Sono pressappoco uguali (coeff. di var.: $\approx 0{,}044$ e $\approx 0{,}045$), con una leggerissima prevalenza per il piede.

4) I due coefficienti di variazione, soprattutto; e anche i due campi di variabilità, rapportati alle rispettive medie.

5) *media* $\approx 20{,}85$; *sc. q. m.* $\approx 16{,}44$; *media per classi di 3 anni* $\approx 20{,}92$; *media per cl. di 5 anni* $\approx 20{,}69$

6) Il più "regolare" è lo studente C che, a parità di media, ha avuto scarto quadratico medio inferiore

7) I padri, il cui coefficiente di variazione è maggiore, sono stati i più disomogenei

11) Ad esempio una catena dimostrativa potrebbe essere la seguente:

$$\sqrt{\frac{(x_1-M)^2+(x_2-M)^2}{2}}=\sqrt{\frac{x_1^2-2Mx_1+M^2+x_2^2-2Mx_2+M^2}{2}}=\sqrt{\frac{x_1^2+x_2^2}{2}+\frac{2M^2-2Mx_1-2Mx_2}{2}}=$$
$$=\sqrt{\frac{x_1^2+x_2^2}{2}+M^2-Mx_1-Mx_2}=\sqrt{\frac{x_1^2+x_2^2}{2}+M^2-M\left(x_1+x_2\right)}=\sqrt{\frac{x_1^2+x_2^2}{2}+M^2-M\cdot 2M}=\sqrt{\frac{x_1^2+x_2^2}{2}-M^2}$$

RISPOSTE agli esercizi di pag. 416-417 (ERRORI DI MISURA)

1) a) $\frac{0+1+2+2+5}{5}=2$ b) $\frac{5-0}{2}=2{,}5$ c) $\frac{|0-2|+|1-2|+|2-2|+|2-2|+|5-2|}{5}=\frac{2+1+0+0+3}{5}=\frac{6}{5}=1{,}2$

d) $\frac{(0-2)^2+(1-2)^2+(2-2)^2+(2-2)^2+(5-2)^2}{5}=\frac{4+1+0+0+9}{5}=\frac{14}{5}=2{,}8$ e) $\sqrt{2{,}8}\approx 1{,}7$

2) a) V

b) F. Scrivendo $G=\overline{x}\pm s$, con $\overline{x}$ media e s scarto quadratico medio,
intendo che pressappoco il 68% delle misure effettuate rientra in quell'intervallo.

Detto n il numero delle misure eseguite, è invece l'intervallo (molto più piccolo) di estremi $\overline{x}\pm\frac{s}{\sqrt{n}}$,
quello che ha circa il 68% di probabilità di contenere il vero valore della grandezza in esame

c) F (*probabilmente*, non "certamente") d) V e) V f) F g) F: *dev.st.pop*() h) F: *quadruplicarlo*

3) a) da $152{,}4-2\cdot 2{,}5=152{,}4-5=147{,}4$ a $152{,}4+2\cdot 2{,}5=152{,}4+5=157{,}4$

b) fra $152{,}4-2\cdot\frac{2{,}5}{\sqrt{625}}=152{,}4-2\cdot\frac{2{,}5}{25}=152{,}4-2\cdot 0{,}1=152{,}4-0{,}2=152{,}2$ e $152{,}4+0{,}2=152{,}6$

4) a) Tra 171,2 e 175,8 b) Arrotondando $2{,}3/\sqrt{64}=0{,}2875$ a $0{,}3$, si ottiene l'intervallo da 173,2 a 173,8

5) a) $\overline{x}\approx 42{,}7$; $s\approx 0{,}9$. Quindi scriveremo $(42{,}7\pm 0{,}9)$ *secondi*.
Il significato della scrittura è che circa il 68% delle misure dovrebbe essere compreso tra $42{,}7-0{,}9$ e $42{,}7+0{,}9$ (estremi esclusi). Se vai a contare il numero di valori nell'intervallo $(42{,}7-0{,}9\ ;\ 42{,}7+0{,}9)$ ossia $(41{,}8\ ;\ 43{,}6)$ ne troverai 53; ora, $53/80=0{,}6625$ che è prossimo a 0,68 (68%) in accordo con quanto detto. Invece i valori tra $\overline{x}-2s$ e $\overline{x}+2s$, ossia tra $42{,}7-1{,}8=40{,}9$ e $42{,}7+1{,}8=44{,}5$ sono 77 e $77/80=0{,}9625$ che è vicino al 95% (0,95) della teoria.

b) Il valore di $s/\sqrt{n}=s/\sqrt{80}$ è $\approx 0{,}1$. Allora un intervallo di confidenza al 95% per il valore della grandezza è quello di estremi $42{,}7\pm 2\cdot 0{,}1=42{,}7\pm 0{,}2$. Ciò significa che tale intervallo $(42{,}5\ ;\ 42{,}9)$ ha una probabilità del 95% circa di contenere il valore sconosciuto del tempo di caduta in esame.

7) da 43 a 45 anni; da 24 a 64 anni 8) Circa il $(100\%-68\%)/2=16\%$

9) Un intervallo che ha il 95 % di probabilità di contenere l'altezza media di tutti i ragazzi di quell'età, residenti in quella città, è quello che va da cm $167-2\cdot 10/\sqrt{420}\approx 166$ a cm $167+2\cdot 10/\sqrt{420}\approx 168$

10) fra 70 e 130 11) $16=10+2\cdot 3$ quindi la distanza dalla media è di 2 dev. st.: circa il $(100\%-95\%)/2=2{,}5\%$

12) Quando le misure sono poche, si utilizza preferibilmente la semidispersione d. In questo caso, $d=1{,}5$ mentre la media delle misure è 86,2; si scriverà il valore della grandezza come $\overline{x}\pm d=86{,}2\pm 1{,}5$

13) a) Si ottiene $2{,}58\pm 0{,}03$ con un arrotondamento ai centesimi per $\overline{x}$

b) La scrittura $2{,}58\pm 0{,}03$ che utilizza la semidispersione dà un'informazione di facile leggibilità sulla media delle misure rilevate e sull'intervallo nel quale approssimativamente si sono distribuite, ma osserviamo che comunque IN GENERE NON TUTTE le misure rientrano nell'intervallo così determinato.
In questo esempio, la misura più piccola è esterna all'intervallo; nel precedente, lo era la misura più grande.

14) La 1^a non è corretta perché non viene specificato di che tipo è l'incertezza; nella 2^a manca l'unità di misura.

15) 1^a scrittura : $incertezza\ rel.=\frac{0{,}04}{3{,}24}=\frac{1}{81}$; 2^a scrittura : $incert.\ rel.=\frac{0{,}5}{40{,}5}=\frac{1}{81}$. Sono precise allo stesso modo!

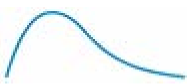
16) $0{,}2/8{,}0=0{,}025$. L'incertezza relativa percentuale è del 2,5%

17) I due "coefficienti di variazione" (rapporti fra scarto quadratico medio e media). E' minore quello del caso a).

18) $2{,}0/\sqrt{n}=0{,}1$; $\sqrt{n}=20$; $n=400$

19) Le rispettive incertezze relative sono: 0,04; 0,03; 0,025. La misura più precisa è dunque la c).

20) La differenza fra 4,97 (media) e 4,60 è 0,37, vicina al doppio dello scarto quadratico medio che è circa 0,18.
Fermo restando che il campione di bustine esaminate è un po' piccolino ai fini di una stima attendibile delle condizioni di tutto l'insieme delle bustine prodotte in una giornata, possiamo presumere che sulle ≈ 48000 peserà non più di 4,6 grammi una percentuale prossima alla metà del 5%, che corrisponde a 1200 bustine.

21) a) Distribuzione con "coda verso destra"  (*positively skewed*) ⇨ b) $media\approx 7{,}9$; $mediana=8$; $moda=9$

RISPOSTE agli esercizi delle pagg. 420-421 (ARROTONDAMENTI E CIFRE SIGNIFICATIVE)

1) a) NO. $x=27{,}9\pm0{,}4$ b) NO. $x=35{,}7\pm0{,}4$ c) NO. $x=2{,}30\pm0{,}05$ d) SI'
e) NO. $x=7{,}34\pm0{,}08$ f) NO. $x=4530\pm50=(453\pm5)\cdot10$ g) NO. $x=91\pm2$ h) SI'

2) a)
$x+y=\cancel{11{,}87}\ \ 11{,}9$
$x-y=\cancel{7{,}13}\ \ 7{,}1$
$xy=\cancel{22{,}515}\ \ 23$
$x/y=\cancel{4{,}008...}\ \ 4{,}0$

b)
$x+y=\cancel{75{,}6}\ \ 76$
$x-y=\cancel{-10{,}4}\ \ -10$
$xy=\cancel{1401{,}8}\ \ \cancel{1400}\ \ 14\cdot10^2$
$x/y=\cancel{0{,}758...}\ \ 0{,}76$

c)
$x+y=\cancel{54{,}54}\ \ 54{,}5$
$x-y=\cancel{54{,}06}\ \ 54{,}1$
$xy=\cancel{13{,}032}\ \ 13$
$x/y=\cancel{226{,}25}\ \ \cancel{230}\ \ 23\cdot10$

d)
$x+y=\cancel{5{,}555}\ \ 5{,}6$
$x-y=\cancel{-0{,}845}\ \ -0{,}8$
$xy=\cancel{7{,}536}\ \ 7{,}5$
$x/y=\cancel{0{,}73593...}\ \ 0{,}74$

3) a) I) assoluta: $\cancel{0{,}53}\ \ 0{,}5$ relativa: $\cancel{0{,}04465...}\ \ 0{,}04$ II) assoluta: $\cancel{0{,}53}\ \ 0{,}5$ relativa: $\cancel{0{,}07433...}\ \ 0{,}07$
III) relativa: $\cancel{0{,}06528...}\ \ 0{,}07$ assoluta: $\cancel{1{,}47}\ \ 1{,}5$ IV) relativa: $\cancel{0{,}06528...}\ \ 0{,}07$ assoluta: $\cancel{0{,}2617...}\ \ 0{,}3$
V) rel.: $\cancel{0{,}21052...}\ \ 0{,}2$ ass.: $\cancel{1714{,}75}\ \ 1700=1{,}7\cdot10^3$ VI) rel.: $\cancel{0{,}02631...}\ \ 0{,}03$ ass.: $\cancel{0{,}08111...}\ \ 0{,}08$

b) I) assoluta: $1{,}4$ (NOTA 1); relativa: $\cancel{0{,}01851...}\ \ 0{,}02$ (NOTA 2)

NOTA 1
Abbiamo detto che nelle scienze sperimentali si solito si osserva la prassi seguente:

L'incertezza Δx viene sempre arrotondata in modo che conservi una cifra significativa soltanto O AL MASSIMO DUE CIFRE SIGNIFICATIVE SE LA PRIMA DI ESSE È 1

Avevamo poi specificato che non tutti sono concordi in questo.
Alcuni accettano nell'incertezza fino a due cifre significative; altri suggeriscono di usare due cifre significative se la prima cifra è bassa (c'è chi dice 1 o 2, c'è chi dice 1, 2, 3 o 4), altrimenti una.

In effetti, se la prima cifra è piccola, eliminare con l'arrotondamento la seconda porterebbe ad una perdita di precisione ritenuta eccessiva anche per un'incertezza.

Vediamo di spiegarci con un esempio.
Se arrotondo $8{,}4$ a 8, come si deforma il mio valore?
Di poco, perché cambia di 0,4; e $0{,}4/8=0{,}05$: cambia quindi del 5%.

Se invece arrotondo $1{,}4$ a 1, qual è la perdita in precisione? E' $0{,}4/1=0{,}4$ che corrisponde addirittura al 40%.

Ecco perché **se la prima cifra è piccola**
(noi abbiamo scelto di considerare tale solo la cifra 1,
altri fanno rientrare nelle cifre "basse" anche il 2, qualcuno si spinge fino al 3 e al 4)
è ragionevole mantenere 2 cifre significative:
la compattezza del dato ne risente un poco, ma si evita una perdita di precisione "importante" in percentuale.

NOTA 2
E' vero che la prima cifra significativa dell'incertezza relativa comincia qui con 1,
e che in questo caso avevamo scritto di tenere due cifre significative anziché una,
ma di fronte al valore $0{,}01851...$ non sembra comunque opportuno fare questa scelta (che porterebbe a 0,019),
perché con l'arrotondamento a 0,02 alteriamo di ben poco, in percentuale, il numero 0,01851…
e in compenso otteniamo una leggibilità decisamente maggiore.

In generale **si incoraggia a usare il "buon senso" in queste scelte se arrotondare o no**, badando,
- ♪ da una parte, che il valore arrotondato non sia molto diverso, in percentuale, rispetto al valore originario,
- ♫ dall'altra alla compattezza e facile leggibilità dell'espressione

e **tenendo sempre presente il contesto**:
- ❑ in che modo sono stati rilevati i dati sperimentali?
- ❑ di che tipo è l'incertezza?
- ❑ che finalità ha il nostro studio, o a chi è rivolta la nostra esposizione?

II) assoluta: $1{,}4$ relativa: $\cancel{0{,}13461...}\ \ 0{,}13$ III) relativa: $\cancel{0{,}03552...}\ \ 0{,}04$ assoluta: $\cancel{49{,}8}\ \ 50=5\cdot10$
IV) relativa: $\cancel{0{,}03552...}\ \ 0{,}04$ assoluta: $\cancel{0{,}02693...}\ \ 0{,}03$
V) rel.: $\cancel{0{,}04907...}\ \ 0{,}05$ ass.: $\cancel{55433{,}5616}\ \ \cancel{60000}\ \ 6\cdot10^4$ VI) rel.: $\cancel{0{,}00613...}\ \ 0{,}006$ ass.: $\cancel{0{,}03502...}\ \ 0{,}04$

c) I) assoluta: $\cancel{0{,}82}\ \ 0{,}8$ relativa: $\cancel{0{,}01503...}\ \ 0{,}015$ II) assoluta: $\cancel{0{,}82}\ \ 0{,}8$ relativa: $\cancel{0{,}01516...}\ \ 0{,}015$
III) relativa: $\cancel{0{,}098...}\ \ 0{,}1$ assoluta: $\cancel{1{,}278}\ \ 1{,}3$ IV) relativa: $\cancel{0{,}098...}\ \ 0{,}1$ assoluta: $\cancel{22{,}1875}\ \ 20=2\cdot10$
V) rel.: $\cancel{0{,}05893...}\ \ 0{,}06$ ass.: $\cancel{512329{,}6224}\ \ \cancel{500000}\ \ 5\cdot10^5$ VI) rel.: $\cancel{0{,}00736...}\ \ 0{,}007$ ass.: $\cancel{0{,}05428...}\ \ 0{,}05$

d) I) assoluta: $\cancel{0{,}106}\ \ 0{,}11$ relativa: $\cancel{0{,}01908...}\ \ 0{,}02$ II) assoluta: $\cancel{0{,}106}\ \ 0{,}11$ relativa: $\cancel{0{,}12544...}\ \ 0{,}13$
III) relativa: $\cancel{0{,}03379...}\ \ 0{,}03$ assoluta: $\cancel{0{,}2547}\ \ 0{,}3$ IV) relativa: $\cancel{0{,}03379...}\ \ 0{,}03$ assoluta: $\cancel{0{,}02487...}\ \ 0{,}02$
V) relativa: $\cancel{0{,}01019...}\ \ 0{,}01$ assoluta: $\cancel{0{,}31346...}\ \ 0{,}3$ VI) rel.: $\cancel{0{,}00127...}\ \ 0{,}0013$ ass.: $\cancel{0{,}00195...}\ \ 0{,}002$